PHAROH lncRNA regulates Myc translation in hepatocellular carcinoma via sequestering TIAR

Allen T. Yu[1,2], Carmen Berasain[3,4,5], Sonam Bhatia[1], Keith Rivera[6], Bodu Liu[1], Frank Rigo[7], Darryl J. Pappin[1], and David L. Spector[1,2]*

[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724
[2]Genetics Program, Stony Brook University, Stony Brook, NY 11794
[3]Hepatology Program, Cima, University of Navarra, Pamplona, Spain
[4]Instituto de Investigaciones Sanitarias de Navarra-IdiSNA, Pamplona, Spain
[5]CIBERehd, Instituto de Salud Carlos III, Madrid, Spain
[6]Broad Institute of MIT and Harvard, Cambridge, MA 02142
[7]Ionis Pharmaceuticals, Carlsbad, CA 92010

*Corresponding author: spector@cshl.edu

Summary

Hepatocellular carcinoma, the most common type of liver malignancy, is one of the most lethal forms of cancer. We identified a long non-coding RNA, *Gm19705*, that is over-expressed in hepatocellular carcinoma and mouse embryonic stem cells. We named this RNA *Pluripotency and Hepatocyte Associated RNA Overexpressed in HCC*, or *PHAROH*. Depletion of *PHAROH* impacts cell proliferation and migration, which can be rescued by ectopic expression of *PHAROH*. RNA-seq analysis of *PHAROH* knockouts revealed that a large number of genes with decreased expression contain a *Myc* motif in their promoter. MYC is decreased at the protein level, but not the mRNA level. RNA-antisense pulldown identified nucleolysin TIAR, a translational repressor, to bind to a 71-nt hairpin within *PHAROH*, sequestration of which increases MYC translation. In summary, our data suggest that *PHAROH* regulates MYC translation by sequestering TIAR and as such represents a potentially exciting diagnostic or therapeutic target in hepatocellular carcinoma.

Introduction

Hepatocellular carcinoma (HCC), the most common type of liver malignancy, is one of the most lethal forms of cancer (Asrani et al., 2019). HCC is the fifth-most frequently diagnosed cancer and the third-leading cause of cancer-related deaths worldwide (Villanueva, 2019). The molecular landscape of HCC is very complex and includes multiple genetic and epigenetic modifications which could represent new diagnosis and therapeutic targets. In this sense, multiple studies have established molecular classifications of HCC subtypes that could be related to clinical management and outcomes (Dhanasekaran et al., 2019; Llovet et al., 2018). For instance, Hoshida et al. classified HCC into S1, S2, and S3 subtypes by means of their histological, pathological, and molecular signatures (Hoshida et al., 2009). S1 tumors exhibit high TGF-β and Wnt signaling activity but do not harbor mutations or genomic changes. The tumors are relatively large, poorly-differentiated, and associated with poor survival. S2 tumors have increased levels of Myc and phospho-Akt and overexpress α-fetoprotein, an HCC serum biomarker. S3 tumors harbor mutations in *CTNNB1* (β-catenin) but tend to be well-differentiated and are associated with good overall survival.

The standard of care for advanced HCC is treatment with sorafenib, a multi-kinase inhibitor that targets Raf, receptor tyrosine kinases (RTKs), and the platelet-derived growth factor receptor (PDGFR). Sorafenib extends the median survival time from 7.9 months to 10.7 months, and lenvatinib, a multiple VEGFR kinase inhibitor, has been reported to extend survival to 13.6 months (Llovet et al., 2018; Philip et al., 2005; Rimassa & Santoro, 2009). Combination therapies of VEGF antagonists together with sorafenib or erlontinib are currently being tested (Dhanasekaran et al., 2019; Greten et al., 2019; Quintela-Fandino et al., 2010). However, even with the most advanced forms

of treatment, the global death toll per year reaches 700,000, creating a mortality ratio of 1.07 with a 5-year survival rate of 18% (Ferlay et al., 2010; Siegel et al., 2014; Villanueva, 2019). Not only is it difficult to diagnose HCC in the early stages, but there is also a poor response to the currently available treatments. Thus, novel therapeutic targets and treatments for HCC are urgently needed.

The ENCODE consortium revealed that as much as 80% of the human genome can be transcribed, while only 2% of the genome encodes for proteins (Djebali et al., 2012). Thousands of transcripts from 200 nucleotides (nt) to over one-hundred kilobases (kb) in length, called long non-coding RNAs (lncRNAs), are the largest and most diverse class of non-protein-coding transcripts. They commonly originate from intergenic regions or introns and can be transcribed in the sense or anti-sense direction. Most are produced by RNA polymerase II and can be capped, spliced and poly-adenylated (reviewed in Rinn & Chang, 2012). Strikingly, many are expressed in a cell or tissue-specific manner and undergo changes in expression level during cellular differentiation and in cancers (Costa, 2005; Dinger et al., 2008). These lncRNAs present as an exciting class of regulatory molecules to pursue, as some are dysregulated in HCC and have potential to be specific to a subtype of HCC (Li et al., 2015).

One of the few examples of a lncRNA that has been studied in the context of HCC is the homeobox (HOX) anti-sense intergenic RNA (*HOTAIR*). This transcript acts in trans by recruiting the Polycomb repressive complex 2 (PRC2), the lysine-specific histone demethylase (LSD1) and the CoREST/REST H3K4 demethylase complex to their target genes (Ezponda & Licht, 2014). *HOTAIR* promotes HCC cell migration and invasion by repressing RNA binding motif protein 38 (RBM38), which is otherwise targeted by p53 to induce cell cycle arrest in G1 (Shu et al., 2006; Yu et al., 2015).

Another mechanism through which lncRNAs function involves inhibitory sequestration of miRNAs and transcription factors (Cesana et al., 2011). In HCC, the lncRNA *HULC* (highly upregulated in liver cancer) sequesters *miR-372*, which represses the protein kinase PRKACB, and down-regulates the tumor suppressor gene *CDKN2C* (p18) (J. Wang et al., 2010). Similarly, the highly-conserved *MALAT1* lncRNA controls expression of a set of genes associated with cell proliferation and migration and is upregulated in many solid carcinomas (Amodio et al., 2018; R. Lin et al., 2007); siRNA knockdown of *MALAT1* in HCC cell lines decreases cell proliferation, migration, and invasion (Lai et al., 2012).

Only a small number of the thousands of lncRNAs have been characterized in regard to HCC. Therefore, whether and how additional lncRNAs contribute to HCC remains unknown, and it is not fully understood how lncRNAs acquire specificity in their mode of action at individual gene loci. A lack of targetable molecules limits the effectiveness of treatments for HCC, and this class of regulatory RNAs has great potential to provide novel therapeutic targets.

Here, we reanalyzed naïve and differentiated transcriptomes of mouse embryonic stem cells (ESCs) in the context of the GENCODE M20 annotation. We aimed to identify lncRNAs that are required for the pluripotency gene expression program, and dysregulated in cancer, with a specific focus on HCC. Since normal development and differentiation are tightly regulated, dysfunction of potential regulatory RNAs may lead to various disease phenotypes including cancer. One lncRNA that is highly upregulated in HCC is of special interest, and we show that it interacts with and sequesters the translation repressor nucleolysin TIAR resulting in an increase of Myc translation. Together, our findings identified a mechanism by which a lncRNA regulates

translation of MYC in HCC by sequestering a translation inhibitor and as such has potential as a therapeutic target in HCC.

Results

Deep sequencing identifies 40 long non-coding RNAs dysregulated in embryonic stem cells and cancer

Since normal development and differentiation are tightly regulated processes, we reasoned that lncRNAs whose expressions are ESC specific and can be found to also exhibit altered expression in cancer, may have important potential roles in regulating critical cellular processes.

We re-analyzed the raw data from our published differential RNA-seq screen comparing lncRNA expression in mouse ESCs vs neural progenitor cells (NPCs) (Bergmann et al., 2015), using updated bioinformatic tools and the recently released GENCODE M20 annotation (January 2019), which has nearly 2.5 times more annotated lncRNAs than the previously used GENCODE M3. Principal component analysis (PCA) of the processed data showed that ESCs and NPCs independently cluster, and the difference between ESC cell lines (AB2.2) and mouse derived ESCs only accounted for 4% of the variance (Figure 1A). Additionally, we prioritized transcripts with an FPKM value greater than 1, and those that were more than 2-fold upregulated in ESCs compared to NPCs. This left us with 147 ESC specific transcripts. Since our goal is to discover novel transcripts that may play a role in the progression of human cancer, we first needed to identify the human homologues of the 147 mouse ESC transcripts. In addition to sequence conservation, we also evaluated syntenic conservation of the mouse lncRNAs to the human genome, due to the fact that many lncRNAs are not

conserved on the sequence level. Finally, we queried TCGA databases via cBioportal, to find lncRNAs that were altered in cancer (Figure 1B). A final candidate list of 40 lncRNAs that are enriched in ESCs, and dysregulated in cancer, was identified (Table 1). Our candidate list contains lncRNAs that have a wide range of expression, and also contains several previously identified lncRNAs that have been found to be dysregulated in cancer (*NEAT1*, *FIRRE*, *XIST*, *DANCR*, and *GAS5*), verifying the validity of the approach (Figure 1–figure supplement 1A) (Ji et al., 2019; Soudyab et al., 2016; Yuan et al., 2016).

We analyzed the ENCODE expression datasets of adult mouse tissue to compare the expression levels of the candidates across tissues (Figure 1C). LncRNAs are known to have distinct expression patterns across different tissues, and our results support the notion that lncRNAs are generally not pan-expressed. Interestingly, many of the identified lncRNAs are enriched in embryonic liver, which is the organ with the most regenerative capacity, yet never grows past its original size.

From here, we decided to focus on liver enriched candidate mouse lncRNAs, especially those that were primarily dysregulated in liver cancers. Because HCC is one of the deadliest cancers and has inadequate treatment options, we focused on lncRNAs that were dysregulated in HCC, *LINC00862*, *TSPOAP-AS1*, *MIR17HG*, and *SNHG5*, with their mouse counterparts being *Gm19705*, *Mir142hg*, *Mir17hg*, and *Snhg5*, respectively. Out of these four lncRNAs that were detected to be amplified in HCC, *LINC00862* was the highest at 13% of all liver cancer cases (Figure 1–figure supplement 1B). We assayed *LINC00862* expression in human samples obtained from healthy and cirrhotic livers and HCC nodules. Indeed, we found that levels of *LINC00862* were elevated in HCC tumor nodules, but also in cirrhotic liver, suggesting

that it may play a role in HCC progression (Figure 1D). In addition, we also assayed *LINC00862* expression in human HCC cell lines and we found it to be upregulated in numerous HCC cell lines compared to the normal human liver cell line, THLE-2 (Figure 1E).

In order to use a more tractable model system, we assessed the conservation of *LINC00862* and its potential mouse counterpart, *GM19705*, which was internally designated as *lnc05* in previous analyses (Bergmann et al., 2015). While much shorter, *GM19705* has 51% sequence identity and the gene order is syntenically conserved, although a reversal event most likely occurred within the locus (Figure 1–figure supplement 1C). Weighted gene correlation network analysis of *GM19705* identified that its expression is highly correlated with those of cell cycle genes, such as *BRCA1* and *BRCA2* (Figure 1–figure supplement 1D). GO-term analysis of the cluster identified cell cycle processes as highly enriched, indicating that *GM19705* may play a role in the regulation of the cell cycle (Figure 1–figure supplement 1E). Re-analysis of previously published single cell analysis of normal adult mouse liver (Tabula Muris et al., 2018) identified *GM19705* expression to be low overall, as expected, but highly expressed exclusively in a subset of hepatocytes (Figure 1–figure supplement 1F).

Our analysis identified *GM19705*/*LINC00862* as a lncRNA that is expressed in ESCs and dysregulated in HCC. We found that *GM19705* is also highly expressed in developing liver and exclusively in adult hepatocytes, and it may have a potential function to regulate the cell cycle. Therefore, we named this mouse lncRNA – *Pluripotency and Hepatocyte Associated RNA Overexpressed in HCC*, or *PHAROH*.

194 *PHAROH is a novel lncRNA that is highly expressed in embryonic liver and mouse*
195 *hepatocellular carcinoma*

196 *PHAROH* is an intergenic lncRNA located on mouse chr1:1qE4. 5' and 3' rapid
197 extension of cDNA ends (RACE) revealed the presence of two isoforms that share two
198 common exons and are both ~450 nt (Figure 2A). In silico analysis of the coding
199 potential by three independent algorithms, which use codon bias (CPAT/CPC) and
200 comparative genomics (PhyloCSF), all point towards the low coding potential score of
201 *PHAROH*, compared to the *Gapdh* control (Figure 2–figure supplement 2A-B). From
202 here on, only qPCR primers that amplify common exons were used. We confirmed
203 expression levels of *PHAROH* in developing liver by assaying the liver bud from E14
204 and E18 embryos and found that they were 7-9 fold enriched compared to adult liver
205 (Figure 2B). Because the liver is one of the main sites of hematopoiesis in the embryo,
206 we measured *PHAROH* levels in embryonic blood and found that expression was
207 exclusive to the liver, and not to hematopoietic cells (Figure 2–figure supplement 2C).
208 *PHAROH* was also found to be upregulated in a partial hepatectomy model of liver
209 regeneration (Figure 2–figure supplement 2D), where the expression was correlated
210 with time points of concerted DNA synthesis, but did not fluctuate across the cell cycle
211 (Figure 2–figure supplement 2E). To confirm *PHAROH*'s involvement in HCC, we used
212 a diethylnitrosamine (DEN) induced carcinogenic model of liver injury. By 11 months
213 post DEN treatment, we were able to visualize HCC tumor nodules, which had elevated
214 levels of *PHAROH* (Figure 2C). In order to facilitate the molecular and biochemical
215 study of *PHAROH*, we chose two mouse HCC cell lines, Hepa1-6 and Hepa1c1c7, and
216 indeed found that *PHAROH* was 3-4-fold more enriched than in ESCs, and 8-10-fold
217 increased over the AML12 mouse normal hepatocyte cell line (Figure 2D).

Single molecule RNA-FISH revealed that *PHAROH* is entirely nuclear in ESCs, with an average of 3-5 foci per cell, whereas it is evenly distributed between the nucleus and cytoplasm in Hepa1-6 cells, with an average of 25 foci per cell (Figure 2E-F). Isoform 1 is expressed mostly in ESCs while isoform 2 of *PHAROH* dominates HCC cell lines (Figure 2A, Figure 2–figure supplement 2F). Cellular fractionation of Hepa1-6 cells corroborates the RNA-FISH determined localization of *PHAROH* as well, which *GAPDH* and *MALAT1* localized correctly to previously determined cellular fractions (Figure 2G). Additional lncRNAs tested, such as *XIST*, *FIRRE*, and *NEAT1*, also localized to their expected cellular fractions (Figure 2–figure supplement 2G). It was also determined that *PHAROH* has a relatively longer half-life in the Hepa1-6 cell line (10.8 h), compared to *MALAT1* (8.0 h), and *XIST* (4.2 h) (Figure 2–figure supplement 2H) (Tani et al., 2012; Yamada et al., 2015). Taken together, *PHAROH* is an embryonic stem cell and fetal liver specific lncRNA, that is upregulated in the context of hepatocellular carcinoma.

Targeted knockout of PHAROH

To evaluate the functional role of *PHAROH*, we generated targeted knockouts using CRISPR/Cas9 technology. Two sgRNA guides were designed to delete a region ~700 bp upstream of the TSS, and ~100 bp downstream of the TSS. We chose to transiently express enhanced specificity Cas9 (eSpCas9-1.1) in order to increase specificity, decrease off-target double stranded breaks, and also to avoid stable integration of Cas9 endonuclease due to its transformative potential (Slaymaker et al., 2016). In addition to using two guides targeting *PHAROH*, we used an sgRNA targeting renilla luciferase as a non-targeting control. Each guide was cloned into a separate

fluorescent protein vector (GFP or mCherry) to allow for subsequent selection. Cells were single cell sorted 48h after nucleofection to account for heterogeneity of deletions among a pooled cell population, which may give certain cells a growth advantage. 85% of the cells were GFP+/mCherry+, and we selected four clones for subsequent analysis (Figure 3–figure supplement 3A). All selected clones had the correct homozygous deletion when assayed by genomic PCR (Figure 3A). qRT-PCR indicated that *PHAROH* was knocked down 80-95% (Figure 3B).

We assayed the proliferative state of the *PHAROH* knockout clones and found a decrease in proliferation. The doubling time of the knockout clones increased to 18.2h, compared to the wildtype doubling time of 14.8h, and ectopic expression of PHAROH reduced the doubling time to nearly wild-type levels (Figure 3C). Ectopic expression of *PHAROH* also successfully rescued the proliferation phenotype in the knockout clones, suggesting that *PHAROH* functions in *trans* (Figure 3D). Migration distance was also decreased by 50% in the knockout clones (Figure 3E).

In addition to assessing the role of *PHAROH* in knockout clones we also employed the use of antisense oligonucleotides (ASO) to knockdown *PHAROH*. We treated cells independently with a control scrambled cEt ASO, or two independent cEt ASOs complementary to the last exon of *PHAROH*. ASOs were nucleofected at a concentration of 2 uM, and we are able to achieve a >90% knockdown at 24h, and a ~50% knockdown was still achieved after 96h (Figure 3–figure supplement 3B). Proliferation assays using manual cell counts and MTS assay shows a 50% reduction in proliferation at 4 days (96h), similar to that achieved in our knockout clones, further supporting a role of *PHAROH* in cell proliferation (Figure 3–figure supplement 3C). Addition of the ASO into the medium allowed for the knockdown to persist for longer

duration to study the impact on clonogenic ability (Figure 3F). Colony formation assays demonstrated that knockdown of *PHAROH* significantly inhibits clonogenic growth of HCC cells in a dose dependent manner (Figure 3G, Figure 3–figure supplement 3D).

To investigate the global effect of *PHAROH* depletion, we performed poly(A)+ RNA-seq on control and knockout clones (Figure 4–figure supplement 4A-B). We identified 810 differentially expressed genes, and GO term analysis revealed regulation of cell proliferation, locomotion, and cell motility as the highest enriched terms (Figure 4A). To determine if these differentially expressed genes were predominantly controlled by common transcription factors, we performed de novo and known motif analysis. Interestingly, promoter motif analysis of differentially expressed genes revealed enrichment of the Myc motif in our dataset suggesting a subset of the genes were under the transcriptional control of Myc (Figure 4–figure supplement 4C). This was intriguing because Myc is known to regulate cell proliferation, and is highly amplified in nearly half of hepatocellular carcinomas (Zheng et al., 2017). However, Myc expression changes were not detected in our RNA-seq analysis, nor was there any statistically significant change compared to sgRenilla controls when assayed by qRT-PCR (Figure 4B). Strikingly, *MYC* protein levels were substantially decreased in all of the *PHAROH* knockout clones, as detected by western blot and immunofluorescence, suggesting that *PHAROH* regulates *Myc* post-transcriptionally (Figure 4C, Figure 4–figure supplement 4D-E). qRT-PCR of genes downstream of Myc that were identified through our analysis were also significantly downregulated in *PHAROH* knockout clones (Figure 4D). Thus, we suggest that depletion of *PHAROH* decreases *MYC* protein levels, and ultimately cell proliferation.

RAP-MS identifies TIAR as the major interactor of PHAROH

LncRNAs can act as structural scaffolds to promote interaction between protein complexes or to sequester a specific protein (Lee et al., 2016; Tsai et al., 2010). Because modulation of *PHAROH* levels change *Myc* protein levels, but not mRNA levels to a significant degree, we hypothesized that *PHAROH* may be regulating the translation of *MYC* through a protein mediator. In order to search for *PHAROH* interacting proteins, we used a pulldown method adapted from the previously published RNA antisense purification-mass spectrometry (RAP-MS) (McHugh et al., 2015). In lieu of pooling all available antisense capture biotinylated oligonucleotides (oligos), we reasoned that individual oligos may be similarly effective, and can be used as powerful biological replicates. In addition, we would minimize oligo-specific off targets by verifying our results with multiple oligos. To this end, we screened through five 20-mer 3' biotinylated DNA oligos that tiled the length of *PHAROH*, and found that four out of the five oligos pulled down >80% of endogenous *PHAROH*, while the pulldown of a control RNA, *PPIB*, remained low. (Figure 5A, Figure 5–figure supplement 5A).

For elution of *PHAROH*, we tested a range of temperatures and found that the elution efficiency reaches the maximum at 40° C, and thus we used this temperature for further experiments (Figure 5B). The remaining level of *PHAROH* RNA on the beads was the direct inverse of the eluate (Figure 5–figure supplement 5B). We chose *PPIB* as a negative control because it is a housekeeping mRNA that is expressed on the same order of magnitude as *PHAROH*, and is not expected to interact with the same proteins. We screened through ten oligos against *PPIB*, and found only one that pulled *PPIB* down at ~60% efficiency, and eluted at the same temperature as *PHAROH* (Figure 5–figure supplement 5C-D). Off-target RNA pulldown, such as *PHAROH* and *18S* rRNA,

remained minimal when using the oligo antisense to *PPIB* (Figure 5–figure supplement 5C).

To identify proteins that bind to *PHAROH*, we analyzed two independent oligos that target *PHAROH*, and two replicates of *PPIB*, on a single 4-plex iTRAQ (isobaric tag for relative and absolute quantitation) mass spectrometry cassette and identified a total of 690 proteins. By plotting the $\log_2$ enrichment ratio of *PHAROH* hits divided by *PPIB* hits, quadrant I will contain proteins that both oligos against *PHAROH* recognize, and quadrant III will be enriched for proteins that bind specifically to *PPIB*. Quadrant III was enriched for keratins, elongation factors, and ribosomal proteins. Interestingly, the top hit in quadrant I is nucleolysin TIAR (TIAL-1), an RNA-binding protein that controls mRNA translation by binding to AU-rich elements in the 3' UTR of mRNA (Figure 5C, Table 2) (Mazan-Mamczarz et al., 2006). TIAR is present in <10% of all experiments queried on Crapome.org (31/411). Immunoblots for TIAR confirm the mass spectrometry data in that TIAR is specific to *PHAROH* pull-down oligos, and also is eluted at 40° C (Fig 5D). Additional controls that are not complementary to the mouse genome and oligos targeting *PHAROH* also confirm the TIAR hit, and it is reproducible in two independent HCC cell lines (Figure 5E). RNase A treatment of the lysate largely abolished the interaction, which indicates that the interaction is RNA mediated, and not the result of direct binding to the oligo (Figure 5E). Immunoprecipitation of TIAR and subsequent extraction of interacting RNA shows enrichment for *PHAROH* when compared to *PPIB* and IgG control (Figure 5F). Thus, together these data indicate that TIAR is a bona fide interactor of *PHAROH*.

A 71-nt sequence in PHAROH has four TIAR binding sites

A previous study on TIAR has mapped its RNA recognition motif across the transcriptome (Meyer et al., 2018). Analysis of *PHAROH*'s sequence reveals that TIAR binding sites are enriched in the 5' end of the transcript of both isoforms (Figure 6A). To determine if there are any conserved structure within *PHAROH* that mediates this interaction, RNA folding prediction algorithms, mFold and RNAfold, were used. The two strongest TIAR binding sequences (TTTT and ATTT/TTTA) were mapped onto ten outputted predicted structures (Figure 6–figure supplement 6A). Strikingly, four out of the seven binding sites consistently mapped to a hairpin that was conserved throughout all predicted structures. Three of the strongest binding motifs localize to the stem of the hairpin, while one secondary motif resides in a bulge (Figure 6B). These data indicate that the sequence is a highly concentrated site for TIAR binding, and is designed to potentially sequester multiple copies of TIAR.

RNA electromobility shift assay (EMSA) of the hairpin and recombinant human TIAR showed that as TIAR concentration increases, it binds to the *PHAROH* hairpin multiple times (Figure 6C). TIAR has a preference to bind two and four times, rather than once or three times. Densitometry quantification of the remaining free probe shows that TIAR has an approximate dissociation constant of 2 nM, consistent with the literature (Kim et al., 2011) (Figure 6–figure supplement 6B). Addition of an antibody against TIAR creates a supershift, showing that the interaction is specific, while addition of IgG does not. The interaction can be abolished with addition of 20x unlabeled probe as well (Figure 6E, left panel).

To determine if binding of TIAR is specific to the sequence and mapped motifs, we created sequential mutations of the hairpin by changing the non-canonical Watson-Crick base pairs (starred and in red) to canonical ones (Figure 6B). Mutation of the first binding site (m1) slightly reduced specificity of TIAR to the hairpin, but changes the preference of TIAR binding to one and two units (Figure 6E, right panel). Mutation of m2 greatly reduced TIAR association, and only two bands are highly visible (Figure 6E, right panel). However, mutation of three binding sites (m3) did not appreciably change the pattern, as compared to m2, perhaps suggesting that the weaker binding site is only used cooperatively (Figure 6–figure supplement 6C). Mutation of all four binding sites (m4) showed minimal TIAR binding (Figure 6E). Taken together, these data indicate that TIAR binds directly to the 71-nt sequence on *PHAROH*, which can fold into a hairpin, and preferentially binds two or four times.

PHAROH modulates Myc translation by sequestering TIAR

TIAR has been shown to bind to the 3' UTR of mRNAs containing AU-rich elements in order to inhibit their translation (Mazan-Mamczarz et al., 2006). It has also been shown that TIAR binds to the 3' UTR of *Myc* mRNA (Liao et al., 2007). Our data suggests that *PHAROH* serves to competitively sequester TIAR in order to allow for increased *MYC* translation. Thus, knockout or knockdown of *PHAROH* will free additional TIAR molecules to bind to the 3' UTR of *Myc* and inhibit its translation.

We began by determining where TIAR binds to *Myc* mRNA. Mapping PAR-CLIP reads from (Meyer et al., 2018) shows two distinct binding sequences on the human *MYC* mRNA, but only one sequence maps to the mouse genome. The stretch of 53-nt

sequence has three distinct regions that are enriched in poly-uridines, but structural prediction largely places the sequences in a loop formation (Figure 7–figure supplement 7A-B). RNA EMSA of the 53-nt 3' UTR and recombinant TIAR showed preference for a singular binding event, and three events are only seen when the binding reaction is saturated by TIAR (Figure 7A). ASO mediated knockdown of *PHAROH* shows reduction of MYC protein similar to the knockouts, but no change in mRNA levels, or TIAR protein levels (Figure 7B, Figure 7–figure supplement 7C). While mRNAs are generally much more highly expressed than lncRNAs, *Myc* is only 3-fold more expressed than *PHAROH* in HCC cell lines (Figure 7B). In addition, there are multiple TIAR binding sites on *PHAROH*, which increases the feasibility of a competition model (Figure 7B).

Next, we tested this hypothesis *in vitro*, by allowing TIAR to bind to the 53-nt Myc 3' UTR, and titrating increasing amounts of *PHAROH* or the mutant *PHAROH* transcript. The wildtype *PHAROH* hairpin can be seen to compete with *Myc* very effectively at nearly all tested ratios, with near complete competition at 10:1 ratio (Figure 7C). However, the fully mutant *PHAROH* was not able to compete with *Myc* nearly as effectively, and was only seen to be slightly effective at the 10:1 ratio (Figure 7C). This data suggests that the *PHAROH* has the capability to successfully compete with the *Myc* 3' UTR binding site in a sequence dependent manner.

In addition, we cloned the full length *Myc* 3'UTR into a dual luciferase reporter construct in order to test our hypothesis in cells. We found that addition of *PHAROH* does indeed increase the luciferase signal by ~50% in a dose dependent manner while the mutant *PHAROH* did not (Figure 7D, Figure 7–figure supplement 7D).

Given that the knockdown or knockout of *PHAROH* reduces MYC levels due to the release of TIAR, we asked whether MYC protein levels would change in the context of *PHAROH* overexpression. Compared to GFP transfection, overexpression of *PHAROH* increases *MYC* protein levels; however, overexpression of mutant *PHAROH* did not change the protein levels of MYC (Figure 7E). Modulation of *PHAROH* or TIAR levels did not have an effect on *Myc* mRNA levels (Figure 7–figure supplement 7E).

Discussion

Studies of the transcriptome have shed important insights into the potential role of the non-coding RNA portion of the genome in basic biology as well as disease. As such, lncRNAs can serve as biomarkers, tumor suppressors, or oncogenes, and have great potential as therapeutic targets (reviewed in Arun et al., 2018). Here, we identified a lncRNA, *PHAROH*, that is upregulated in mouse ESCs, embryonic and regenerating adult liver and in HCC. It also has a conserved human ortholog, which is upregulated in human patient samples from cirrhotic liver and HCC. Genetic knockout or ASO knockdown of *PHAROH* results in a reduction of cell proliferation, migration, and colony formation.

To elucidate the molecular mechanism through which *PHAROH* acts in proliferation, we used RNA-seq and mass spectrometry to provide evidence that *PHAROH* regulates MYC translation via sequestering the translational repressor TIAR in *trans*. Modulation of *PHAROH* levels reveal that it is positively correlated with MYC protein level, which is well known to be associated with HCC and is amplified in nearly 50% of HCC tumors (Peng et al., 1993). In addition, MYC has been characterized as a

critical player in liver regeneration (Zheng et al., 2017). We identified TIAR as an intermediate player in the *PHAROH*-MYC axis, which has been reported to bind to the 3' UTR of MYC mRNA and suppress its translation (Mazan-Mamczarz et al., 2006). While TIAR is an RNA-binding protein that is known for its role in stress granules (Kedersha et al., 1999), we do not detect stress granule formation in our HCC cell lines as assayed by immunofluorescence for TIAR (Figure 7–figure supplement 7F). As such, the role of *PHAROH*-TIAR lies outside the context of stress granule function. Interestingly, overexpression of TIAR is a negative prognostic marker for HCC survival (Figure 7–figure supplement 7G) (Uhlen et al., 2017). As the primary mutation of HCC is commonly amplification of MYC, it is possible that TIAR is upregulated in an attempt to curb MYC expression.

Our analysis maps the *PHAROH*-TIAR interaction to predominantly occur at a 71-nt hairpin at the 5' end of *PHAROH*. While *PHAROH* has two main isoforms that are selectively expressed in ESCs and HCC, the hairpin is commonly expressed in both isoforms. TIAR has been classified as an ARE binding protein that recognizes U-rich and AU-rich sequences. Kinetic and affinity studies have found that TIAR has a dissociation constant of ~1 nM for U-rich sequences, and ~14 uM for AU-rich sequences (Kim et al., 2011). One question that is apparent in the RNA-binding protein field is how RBPs acquire their specificity. While there have been studies that analyze target RNA structure or RNA recognition motif structure, why RBPs bind one transcript over another with a similar sequence is still an open question. For example, the 3' UTR of Myc contains multiple U-rich stretches, ranging from 3 to 9 resides. It has been reported that TIAR binds efficiently to uridylate residues of 3-11 length, yet PAR-CLIP data only reveals two binding events in the human *MYC* transcript (Kim et al., 2011). In

addition, the 53-nt fragment that was assayed in this study contained potentially six TIAR binding sites, yet RNA EMSA analysis revealed a preference for a single binding event (Figure 7A). One explanation is that *PHAROH*'s hairpin has uniquely spaced TIAR binding sites. Because the absolute affinity of TIAR to U-rich sequences is relatively high, one molecule may sterically block additional binding events. However, if the binding sites are properly spaced, binding events will be ordered and perhaps even cooperative. The average gap between binding sites in the *Myc* fragment is 2 nt, while it is 10 nt in the *PHAROH* hairpin, which allows more flexibility in spacing between each bound protein.

In addition, one aspect that was not explored was the requirement for the formation of the hairpin for TIAR binding. Previous studies used synthesized linear oligos as substrates to test the kinetics of these RBPs, and we also mutated the hairpin in a way such that structure is preserved. TIAR contains three RNA recognition motifs (RRM), which typically recognizes single stranded RNA. Therefore, binding of TIAR to the 71-nt sequence of *PHAROH* would require unwinding of the potential hairpin, which is energetically unfavorable. It is also known that TIAR's RRM2 mainly mediates ssRNA polyU-binding, but its dsRNA binding capabilities have not been explored (Kim et al., 2013). There are examples where multiple RRMs in tandem can allow for higher RNA binding affinity and possibly sandwiching dsRNA, and thus it would be possible that TIAR binding to the multiple sites on the *PHAROH* hairpin is cooperative (Allain et al., 2000).

While TIAR may be *PHAROH*'s top interacting protein, it is unknown whether *PHAROH* is TIAR's highest interacting RNA. This would depend on the relative abundances of each RNA species that has the potential to bind TIAR, and TIAR's

474 expression level. This seems to be cell type specific, as TIAR was initially studied in
475 immune cells and was shown to predominantly translationally repress *Tnf-α* through
476 binding of the AU-rich sequence in the 3' UTR (Piecyk et al., 2000). In our cell lines,
477 *Tnf-α* is not expressed. Conversely, a screen for proteins that bind to the *Tnf-α* 3' UTR
478 may not necessarily indicate TIAR as a binder, as evidenced by a recent study (Ma &
479 Mayr, 2018). Another recent study had shown that lncRNA *MT1JP* functions as a tumor
480 suppressor and had the capability to bind to TIAR, which suppresses the translation of
481 p53 (Liu et al., 2016). However, *MT1JP* is largely cytoplasmic, while TIAR in our context
482 is mainly nuclear. Thus, while TIAR may bind additional mRNAs or lncRNAs, it seems
483 that one of the main targets in HCC cell lines is *Myc*, as supported by statistically
484 significant promoter enrichment of the downstream targets.

485 In summary, we have identified a lncRNA, *PHAROH*, that is enriched in ESCs
486 and dysregulated in HCC, and found that it acts to sequester TIAR through a hairpin
487 structure in order to regulate *MYC* translation. Additionally, based on synteny and
488 upregulation in human HCC samples, we identified *LINC00862* as the possible human
489 ortholog of PHAROH (Figure 1D). Future studies will reveal the therapeutic potential of
490 targeting *PHAROH* to impact liver development/regeneration and HCC.

Experimental procedures

Cell culture and genomic PCR

All cell culture reagents were obtained from Gibco (Life Technologies), unless stated otherwise. Hepa1-6, Hepa1c1c7, AML12, SNU-182, THLE-2 cells were obtained from ATCC. Huh7, SNU-387, Hep3B, and HepG2 were generous gifts from Scott Lowe (MSKCC). Hepa1-6, Hepa1c1c7, Huh7, Hep3B, and HepG2 were maintained in DMEM supplemented with 10% FBS and 1% penicillin/streptomycin. SNU-182 and SNU-387 were maintained in RPMI supplemented with 10% FBS and 1% penicillin/streptomycin. AML12 was maintained in DMEM:F12 Medium supplemented with 10% fetal bovine serum, 10 µg/ml insulin, 5.5 µg/ml transferrin, 5 ng/ml selenium, and 40 ng/ml dexamethasone. THLE-2 cells were maintained in BEGM (BEGM Bullet Kit; CC3170) where gentamycin/amphotericin and epinephrine were discarded, and extra 5 ng/mL EGF, 70 ng/mL phosphoethanolamine and 10% fetal bovine serum were added in addition to the supplied additives. ESCs and NPCs were cultured as in (Bergmann et al., 2015). All cells were cultured in a humidified incubator at 37° C and 5% CO_2. Half-life of RNA was determined by adding α-amanitin to a final concentration of 5 µg/mL. Genomic DNA was isolated using DNeasy Blood & Tissue (Qiagen). All cell lines were tested for mycoplasma regularly.

Cellular Fractionation

Cellular fractionation was performed according to (https://link.springer.com/protocol/10.1007%2F978-1-4939-4035-6_1). In brief, cells were collected and resuspended in NP-40 lysis buffer. The cell suspension is overlaid on top of a sucrose buffer and centrifuged at 3,500 x g for 10 minutes to pellet the nuclei.

The supernatant (cytoplasm) is collected and the nuclei are resuspended in glycerol buffer and urea buffer is added to separate the nucleoplasm and chromatin. The cells are centrifuged at 14,000 x g for 2 minutes and the supernatant (nucleoplasm) is collected, while the chromatin-RNA is pelleted.

DEN administration

Mice were injected intraperitoneally with diethylnitrosamine (DEN) at 14 days of age as described (Garcia-Irigoyen et al., 2015). DEN-treated mice, and the corresponding controls injected with saline, were sacrificed at 5, 8, and 11 months post injection.

Partial hepatectomy

Two-thirds partial hepatectomy (PH) and control sham operations (SH) were performed as reported (Berasain et al., 2005). Two SH and four PH mice were sacrificed at 3, 6, 24, 48 and 72 hours after surgery. Animal experimental protocols were approved (CEEA 062-16) and performed according to the guidelines of the Ethics Committee for Animal Testing of the University of Navarra.

Human samples

Samples from patients included in the study were provided by the Biobank of the University of Navarra (CEI 47/2015) and were processed following standard operating procedures approved by the Ethical and Scientific Committees. Liver samples from healthy patients were collected from individuals with normal or minimal changes in the liver at surgery of digestive tumors or from percutaneous liver biopsy performed because of mild alterations of liver function. Samples for cirrhotic liver and HCC were obtained from patients undergoing partial hepatectomy and/or liver transplantation.

The biobank obtained an informed consent and consent to publish from each patient and codified samples were provided to the researchers. The study protocol conformed to the ethical guidelines of the 1975 Declaration of Helsinki. Samples were processed following standard operating procedures approved by the Ethical and Scientific Committees. Liver samples from healthy patients were collected from individuals with normal or minimal changes in the liver at surgery of digestive tumors or from percutaneous liver biopsy performed because of mild alterations of liver function. Samples for cirrhotic liver and HCC were obtained from patients undergoing partial hepatectomy and/or liver transplantation.

Immunoblotting

To determine protein levels in our system, we used 10% SDS-PAGE gels. Gels were loaded with 1-µg protein per well (Bradford assay). The following antibodies were used: β-actin (1:15,000; Sigma), c-Myc (1:1000; CST), TIAR (1:1000; Cell Signaling). IRDye-800CW was used as a fluor for secondary anti-rabbit antibodies, and IRDye-680RD was used for mouse secondary antibodies. Blots were scanned using the Li-Cor Odyssey Classic.

Immunoprecipitation

For TIAR immunoprecipitation, one 10 cm plate of Hepa1c1c7 cells at 80% confluence was lysed in 1 mL Pierce IP Lysis Buffer (supplemented with 100 U/mL SUPERase-IN and 1X Roche protease inhibitor cocktail) and incubated on ice for 10 min. Lysates centrifuged at 13,000xg for 10 minutes. 3 ug of TIAR antibody or rabbit IgG were incubated with the lysate at 4°C for 1 hour. 16 uL of Protein A magnetic beads were washed and added to the lysate and incubated for an additional 30 minutes at 4°C.

50% of beads were resuspended in Laemmli buffer for western blotting and RNA was isolated from the remaining beads using TRIzol.

Immunofluorescence staining

#1.5 round glass coverslips were prepared by acid-cleaning prior to seeding cells. Staining was performed as published previously (Spector, D.L. and H.C. Smith. 1986. Exp. Cell Res. 163, 87-94). In brief, cells were fixed in 2% PFA for 15 min, washed with PBS, and permeabilized in 0.2% Triton X-100 plus 1% normal goat serum (NGS). Cells were washed again in PBS+1% NGS, and incubated with TIAR antibody (1:2000; CST) for 1 hour at room temp in a humidified chamber. Cells were washed again PBS+1% NGS, and incubated with Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 488 (1:1000; Thermo Fisher) secondary antibody for 1 hour at room temp. Cover slips were washed with PBS before mounting with ProLong Diamond antifade (Thermo Fisher).

Cell viability assays

Cells were seeded at a density of 10,000 cells/well (100 µl per well) into 24-well plates and treated with 2.5 µM of either a *PHAROH*-specific ASO or scASO. Cells were grown for 96 h at 37°C. 20 µl of solution (CellTiter 96® AQueous One Solution Reagent, Promega) was added to the wells and incubated for 4 h at 37°C. Measurements of absorbance at 490 nm were performed using a SpectraMax i3 Multi-Mode Detection Platform (Molecular Devices). Background absorbance at 690 nm was subtracted. Cells were also trypsinized, pelleted, and manually counting using a hemocytometer.

RNA antisense pulldown and mass spectrometry

RNA antisense pulldown—Cells were lysed on a 10 cm plate in 1 mL IP lysis buffer (IPLB, 25 mM Tris-HCl pH 7.4, 150 mM NaCl, 1% NP-40, 1 mM EDTA, 5% glycerol, supplemented with 100 U/mL SUPERase-IN and 1X Roche protease inhibitor cocktail) for 10 minutes, and lysate was centrifuged at 13,000xg for 10 minutes. Cell lysate was adjusted to 0.3 mg/mL (Bradford assay). 100 pmol of biotinylated oligo was added to 500 uL of lysate and incubated at room temperature for 1 hour with rotation. 100 uL streptavidin Dynabeads were washed in IPLB, added to the lysate, and incubated for 30 minutes at room temperature with rotation. Beads were washed three times with 1 mL lysis buffer. For determining temperature for optimal elution, beads were then resuspended in 240 uL of 100 mM TEAB and aliquoted into eight PCR tubes. Temperature was set on a veriflex PCR block and incubated for 10 minutes. Beads were captured and TRIzol was added to the eluate and beads. Once optimal temperature is established, the beads were resuspended in 90 uL of 100 mM TEAB, and incubated at 40° C for 10 minutes. TRIzol was added to 30 uL of the eluate, another 30 uL was kept for western blots, and the last 30 uL aliquot was sent directly for mass spectrometry.

Tryptic digestion and iTRAQ labeling—Eluted samples were reduced and alkylated with 5 mM DTT and 10 mM iodoacetamide for 30 min at 55 °C, then digested overnight at 37 °C with 1 µg Lys-C (Promega, VA1170) and dried in vacuo. Peptides were then reconstituted in 50 µl of 0.5 M TEAB/70% ethanol and labeled with 4-plex iTRAQ reagent for 1 h at room temperature essentially as previously described (Ross et al., 2004). Labeled samples were then acidified to <pH 4 using formic acid, combined and concentrated in vacuo until ~10 µl remained.

Two-dimensional fractionation—Peptides were fractionated using a Pierce™ High pH Reversed-Phase Peptide Fractionation Kit (Thermo Scientific, 84868) according to the manufacturer's instructions with slight modifications. Briefly, peptides were reconstituted in 150 µl of 0.1% TFA, loaded onto the spin column, and centrifuged at 3000 × g for 2 min. Column was washed with water, and then peptides were eluted with the following percentages of acetonitrile (ACN) in 0.1% triethylamine (TEA): 5%, 7.5%, 10%, 12.5%, 15%, 20%, 30%, and 50%. Each of the 8 fractions was then separately injected into the mass spectrometer using capillary reverse-phase LC at low pH.

Mass spectrometry—An Orbitrap Fusion Lumos mass spectrometer (Thermo Scientific), equipped with a nano-ion spray source was coupled to an EASY-nLC 1200 system (Thermo Scientific). The LC system was configured with a self-pack PicoFrit™ 75-µm analytical column with an 8-µm emitter (New Objective, Woburn, MA) packed to 25 cm with ReproSil-Pur C18-AQ, 1.9 µM material (Dr. Maish GmbH). Mobile phase A consisted of 2% acetonitrile; 0.1% formic acid and mobile phase B consisted of 90% acetonitrile; 0.1% formic acid. Peptides were then separated using the following steps: at a flow rate of 200 nl/min: 2% B to 6% B over 1 min, 6% B to 30% B over 84 min, 30% B to 60% B over 9 min, 60% B to 90% B over 1 min, held at 90% B for 5 min, 90% B to 50% B over 1 min and then flow rate was increased to 500 µl/min as 50% B was held for 9 min. Eluted peptides were directly electrosprayed into the MS with the application of a distal 2.3 kV spray voltage and a capillary temperature of 300 °C. Full-scan mass spectra (Res = 60,000; 400–1600 m/z) were followed by MS/MS using the "Top Speed" method for selection. High-energy collisional dissociation (HCD) was used with the normalized collision energy set to 35 for fragmentation, the isolation width set to 1.2 and

a duration of 15 s was set for the dynamic exclusion with an exclusion mass width of 10 ppm. We used monoisotopic precursor selection for charge states 2+ and greater, and all data were acquired in profile mode.

Database searching

Peaklist files were generated by Proteome Discoverer version 2.2.0.388 (Thermo Scientific). Protein identification was carried out using both Sequest HT (Eng et al., 1994) and Mascot 2.5 (Perkins et al., 1999) against the UniProt mouse reference proteome (57,220 sequences; 26,386,881 residues). Carbamidomethylation of cysteine, iTRAQ4plex (K), and iTRAQ4plex (N-term) were set as fixed modifications, methionine oxidation, and deamidation (NQ) were set as variable modifications. Lys-C was used as a cleavage enzyme with one missed cleavage allowed. Mass tolerance was set at 20 ppm for intact peptide mass and 0.3 Da for fragment ions. Search results were rescored to give a final 1% FDR using a randomized version of the same Uniprot mouse database, with two peptide sequence matches (PSMs) required. iTRAQ ratio calculations were performed using Unique and Razor peptide categories in Proteome Discoverer.

RNA Electromobility shift assay

DNA template used for in vitro synthesis of RNA probes were from annealed oligos. A T7 RNA polymerase promoter sequence was added to allow for in vitro transcription using the MEGAscript T7 transcription kit (Thermo Fisher). RNA was end labelled at the 3' end with biotin using the Pierce RNA 3' End Biotinylation Kit (Thermo Fisher). RNA quantity was assayed by running an RNA 6000 Nano chip on a 2100 Bioanalyzer. Six percent acrylamide gels (39:1 acrylamide:bis) (Bio-Rad) containing 0.5

X TBE were used for all EMSA experiments. Recombinant human TIAR (Proteintech) was added at indicated concentrations to the probe (~2 fmol) in 20 uL binding buffer, consisting of 10 mM HEPES (pH 7.3), 20 mM KCL, 1 mM Mg_2Cl_2, 1 mM DTT, 30 ng/uL BSA, 0.01% NP-40, and 5% glycerol. After incubation at room temperature for 30 minutes, 10 uL of the samples were loaded and run for 1 hr at 100 V. The nucleic acids were then transferred onto a positively charged nylon membrane (Amersham Hybond-N+) in 0.5 X TBE for 30 minutes at 40 mAh. Membranes were crosslinked using a 254 nM bulb at 120 mJ/cm^2 in a Stratalinker 1800. Detection of the biotinylated probe was done using the Chemiluminescent Nucleic Acid Detection Module Kit (Thermo Fisher 89880).

3' UTR luciferase assay

The full length 3' UTR of c-Myc was cloned into the pmirGLO Dual-Luciferase miRNA target expression vector (Promega). Luciferase activity was assayed in transfected cells using the Dual-Luciferase Reporter Assay (Promega). To evaluate the interaction between *PHAROH*, 3' UTR of c-Myc, and TIAR, cells were transfected with the respective constructs using Lipofectamine 3000. Twenty-four hours later, firefly and Renilla luciferase activity was measured, and Renilla activity was used to normalize firefly activity.

Single Molecule RNA FISH

#1.5 round glass coverslips were prepared by acid-cleaning and layered with gelatin for 20 minutes, prior to seeding MEF feeder cells and ESCs. Cells were fixed for 30 minutes in freshly-prepared 4% PFA (Electron Microscopy Sciences), diluted in D-PBS without $CaCl_2$ and $MgCl_2$ (Gibco, Life Technologies) and passed through a 0.45

675 µm sterile filter. Fixed cells were dehydrated and rehydrated through an ethanol
676 gradient (50% - 75% - 100% - 75% - 50%- PBS) prior to permeabilization for 5 minutes
677 in 0.5% Triton X-100. Protease QS treatment was performed at a 1:8,000 dilution.
678 QuantiGene ViewRNA (Affymetrix) probe hybridizations were performed at 40°C in a
679 gravity convection incubator (Precision Scientific), and incubation time of the pre-
680 amplifier was extended to 2 hours. Nuclei were counter-stained with DAPI and
681 coverslips mounted in Prolong Gold anti-face medium
682 (www.spectorlab.labsites.cshl.edu/protocols).

683 Coverslips were imaged on a DeltaVision Core system (Applied Precision),
684 based on an inverted IX-71 microscope stand (Olympus) equipped with a 60x U-
685 PlanApo 1.40 NA oil immersion lens (Olympus). Images were captured at 1x1 binning
686 using a CoolSNAP HQ CCD camera (Photometric) as z-stacks with a 0.2 µm spacing.
687 Stage, shutter and exposure were controlled through SoftWorx (Applied Precision).
688 Image deconvolution was performed in SoftWorx.

689 A spinning-disc confocal system (UltraVIEW Vox; PerkinElmer) using a scanning
690 unit (CSU-X1; Yokogawa Corporation of America) and a charge-coupled device camera
691 (ORCA-R2; Hamamatsu Photonics) fitted to an inverted microscope (Nikon) equipped
692 with a motorized piezoelectric stage (Applied Scientific Instrumentation). Image
693 acquisition was performed using Volocity versions 5 and 6 (PerkinElmer). Routine
694 imaging performed using Plan Apochromat 60 or 100× oil immersion objectives, NA 1.4.

695 *RNA sequencing and analysis*

696 Total RNA was isolated either directly from cryosections of the tumor tissue or
697 from organotypic epithelial cultures using TRIzol according to the manufacturer's

instructions. RNA quality was assayed by running an RNA 6000 Nano chip on a 2100 Bioanalyzer. For high-throughput sequencing, RNA samples were required to have an RNA integrity number (RIN) 9 or above. TruSeq (Illumina) libraries for poly(A)+ RNA-seq were prepared from 0.5–1mg RNA per sample. To ensure efficient cluster generation, an additional gel purification step of the libraries was applied. The libraries were multiplexed (12 libraries per lane) and sequenced single-end 75 bp on the NextSeq500 platform (Illumina), resulting in an average 40 Million reads per library. Analysis was performed in GalaxyProject. In brief, reads were first checked for quality using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and a minimum Phred score of 30 was required. Reads were then mapped to the mouse mm10 genome using STAR (Dobin et al., 2013), and counts were generating using htseq-counts with the appropriate GENCODE M20 annotation. Deseq2 was then used to generate the list of differentially expressed genes (Love et al., 2014). Motif analysis was performed using HOMER (Heinz et al., 2010).

Coding analysis

cDNA sequences of *PHAROH* and *GAPDH* were inputted into CPAT (http://lilab.research.bcm.edu/cpat/) or CPC (http://cpc.cbi.pku.edu.cn/programs/run_cpc.jsp) for analysis (Kong et al., 2007; L. Wang et al., 2013). PhyloCSF analysis was performed using the UCSC Genome Browser track hub (https://data.broadinstitute.org/compbio1/PhyloCSFtracks/trackHub/hub.DOC.html) (M. F. Lin et al., 2011).

Plasmid construction

721 eSpCas9(1.1) was purchased from Addgene (#71814). eSpCas9-2A-GFP was
722 constructed by subcloning 2A-GFP from pSpCas9(BB)-2A-GFP (PX458) (Addgene
723 #48138) into eSpCas9 using EcoRI sites. To construct eSpCas9-2A-mCherry, 2A-
724 mCherry was amplified from mCherry-Pol II (Zhao et al., 2011), and an internal BbsI site
725 was silently mutated. The PCR product was then cloned into eSpCas9 using EcoRI
726 sites. The *PHAROH* construct was amplified using Hepa1-6 cDNA as a template and
727 cloned into pCMV6 using BamHI and FseI. Mutant *PHAROH* was constructed by
728 amplifying tiled oligos and cloned into pCMV6 using BamHI and FseI.

729 *CRISPR/Cas9 genetic knockout*

730 To generate a genetic knockout of *PHAROH*, two sgRNAs targeting the promoter
731 region were combined, creating a deletion including the TSS. Guide design was
732 performed on Benchling (https://benchling.com) taking into account both off-target
733 scores and on-target scores. The sgRNA targeting the gene body of *PHAROH* was
734 cloned into a pSpCas9(BB)-2A-GFP vector (PX458, Addgene plasmid #48138) and the
735 sgRNA targeting the upstream promoter region was cloned into a pSpCas9(BB)- 2A-
736 mCherry vector. Hepa1-6 were transfected with both plasmids using the 4D-
737 Nucleofector™ System (Lonza) using the EH-100 program in SF buffer. To select for
738 cells expressing both gRNAs, GFP and mCherry double positive cells were sorted 48
739 hours post transfection, as single cell deposition into 96-well plates using a FACS Aria
740 (SORP) Cell Sorter (BD). Each single cell clone was propagated and analyzed by
741 genomic PCR and qRT-PCR to select for homozygous knockout clones. Cells
742 transfected with a sgRNA targeting Renilla luciferase were used as a negative control.

743 *Cell cycle analysis*

744 Hoechst 33342 (Sigma) was added to cells at a final concentration of 5 µg/mL
745 and incubated at 37° C for 1 hour. Cells were trypsinized and collected into a flow
746 cytometry compatible tube. Profiles were analyzed using a FACS Aria (SORP) Cell
747 Sorter (BD), gated according to DNA content and cell cycle phase, and sorted into
748 Eppendorf tubes for subsequent RNA extraction and qRT-PCR analysis.

749 *Nucleofection*

750 For transfection of ASOs using nucleofection technology (Lonza), ESCs were
751 harvested following soaking off of feeder cells for one hour, washed in D-PBS (Gibco,
752 Life Technologies) and passed through a 70 µm nylon cell strainer (Corning). Cell count
753 and viability was determined by trypan blue staining on a Countess automated cell
754 counter (Life Technologies). For each reaction, $1x10^6$ viable cells were resuspended in
755 SF Cell Line Solution (Lonza), mixed with 2 µM control or 2 µM target-specific ASO and
756 transferred to nucleocuvettes for nucleofection on a 4D-Nucleofector System (Lonza)
757 using program code “EH-100”. For plasmid nucleofections, 10 ug of plasmid was used
758 and nucleofected using program code “EH-100”. Cells were subsequently transferred
759 onto gelatinized cell culture plates containing pre-warmed and supplemented growth
760 medium. Growth medium was changed once after 16 hours.

761 *Colony Formation Assay*

762 200 Hepa1-6 cells were seeded in a 6-well plate. ASOs were added at the time
763 of seeding at the indicated concentrations. Two weeks later, cells were fixed, stained
764 with Giemsa, counted and photographed.

765 *2’-O-Methoxyethyl (MOE) antisense oligonucleotides and knockdown analysis*

Synthesis and purification of all 2'-MOE modified oligonucleotides was performed as previously described (Meng et al. 2014) by Ionis Pharmaceuticals. These ASOs are 20-mer oligonucleotides containing a phosphorothioate backbone, 2'-O-methoxyethyl modifications on the first and last five nucleotides and a stretch of ten DNA bases in the center. Constrained ethyl oligos are 16-mer oligonucleotides that contain modifications on the first and last 3 nucleotides and a stretch of ten DNA bases in the center.

qRT-PCR

To assess knockdown efficiency TRIzol-extracted RNA was treated with RNAse-free DNAseI (Life Technologies) and subsequently reverse-transcribed into cDNA using TaqMan Reverse Transcription reagents and random hexamer oligonucleotides (Life Technologies). Real-time PCR reactions were prepared using Power SYBR Green Master Mix (Life Technologies) and performed on an ABI StepOnePlus Real-Time PCR system (Life Technologies) for 40 cycles of denaturation at 95°C for 15 seconds followed by annealing and extension at 60°C for 60 seconds. Primers were designed to anneal within an exon to detect both primary and processed transcripts. Primer specificity was monitored by melting curve analysis. For each sample, relative abundance was normalized to the housekeeping gene *PPIB* mRNA levels.

Declaration of Interests

D.L.S. is a consultant to, and receives research support from, Ionis Pharmaceuticals.

Acknowledgments

We thank members of the Spector lab for critical discussions and advice throughout the course of this study. We would also like to thank the CSHL Cancer Center Shared Resources (Microscopy, Mass Spectrometry, Flow Cytometry, and Next-Gen Sequencing) for services and technical expertise (NCI 2P3OCA45508). This research was supported by NCI 5P01CA013106-Project 3 and 5R35GM131833 (D.L.S.) and NCI 5F31CA220997-02 (A.T.Y.)

References

Allain, F. H., Bouvet, P., Dieckmann, T., & Feigon, J. (2000). Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin. EMBO J, 19(24), 6870-6881. doi:10.1093/emboj/19.24.6870

Amodio, N., Raimondi, L., Juli, G., Stamato, M. A., Caracciolo, D., Tagliaferri, P., & Tassone, P. (2018). MALAT1: a druggable long non-coding RNA for targeted anti-cancer approaches. J Hematol Oncol, 11(1), 63. doi:10.1186/s13045-018-0606-4

Arun, G., Diermeier, S. D., & Spector, D. L. (2018). Therapeutic Targeting of Long Non-Coding RNAs in Cancer. Trends Mol Med, 24(3), 257-277. doi:10.1016/j.molmed.2018.01.001

Asrani, S. K., Devarbhavi, H., Eaton, J., & Kamath, P. S. (2019). Burden of liver diseases in the world. J Hepatol, 70(1), 151-171. doi:10.1016/j.jhep.2018.09.014

Berasain, C., Garcia-Trevijano, E. R., Castillo, J., Erroba, E., Lee, D. C., Prieto, J., & Avila, M. A. (2005). Amphiregulin: an early trigger of liver regeneration in mice. Gastroenterology, 128(2), 424-432. doi:10.1053/j.gastro.2004.11.006

Bergmann, J. H., Li, J., Eckersley-Maslin, M. A., Rigo, F., Freier, S. M., & Spector, D. L. (2015). Regulation of the ESC transcriptome by nuclear long noncoding RNAs. Genome Res, 25(9), 1336-1346. doi:10.1101/gr.189027.114

Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M., . . . Bozzoni, I. (2011). A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. Cell, 147(2), 358-369. doi:10.1016/j.cell.2011.09.028

Costa, F. F. (2005). Non-coding RNAs: new players in eukaryotic biology. Gene, 357(2), 83-94. doi:10.1016/j.gene.2005.06.019

819 Dhanasekaran, R., Nault, J. C., Roberts, L. R., & Zucman-Rossi, J. (2019). Genomic Medicine
820 and Implications for Hepatocellular Carcinoma Prevention and Therapy.
821 Gastroenterology, 156(2), 492-509. doi:10.1053/j.gastro.2018.11.001
822 Dinger, M. E., Amaral, P. P., Mercer, T. R., Pang, K. C., Bruce, S. J., Gardiner, B. B., . . .
823 Mattick, J. S. (2008). Long noncoding RNAs in mouse embryonic stem cell pluripotency
824 and differentiation. Genome Res, 18(9), 1433-1445. doi:10.1101/gr.078378.108
825 Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., . . . Gingeras, T. R.
826 (2012). Landscape of transcription in human cells. Nature, 489(7414), 101-108.
827 doi:10.1038/nature11233
828 Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., . . . Gingeras, T. R.
829 (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 29(1), 15-21.
830 doi:10.1093/bioinformatics/bts635
831 Eng, J. K., McCormack, A. L., & Yates, J. R. (1994). An approach to correlate tandem mass
832 spectral data of peptides with amino acid sequences in a protein database. J Am Soc Mass
833 Spectrom, 5(11), 976-989. doi:10.1016/1044-0305(94)80016-2
834 Ezponda, T., & Licht, J. D. (2014). Molecular pathways: deregulation of histone h3 lysine 27
835 methylation in cancer-different paths, same destination. Clin Cancer Res, 20(19), 5001-
836 5008. doi:10.1158/1078-0432.CCR-13-2499
837 Ferlay, J., Shin, H. R., Bray, F., Forman, D., Mathers, C., & Parkin, D. M. (2010). Estimates of
838 worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer, 127(12), 2893-
839 2917. doi:10.1002/ijc.25516
840 Garcia-Irigoyen, O., Latasa, M. U., Carotti, S., Uriarte, I., Elizalde, M., Urtasun, R., . . . Avila, M.
841 A. (2015). Matrix metalloproteinase 10 contributes to hepatocarcinogenesis in a novel

842 crosstalk with the stromal derived factor 1/C-X-C chemokine receptor 4 axis. Hepatology,
843 62(1), 166-178. doi:10.1002/hep.27798
844 Greten, T. F., Lai, C. W., Li, G., & Staveley-O'Carroll, K. F. (2019). Targeted and Immune-
845 Based Therapies for Hepatocellular Carcinoma. Gastroenterology, 156(2), 510-524.
846 doi:10.1053/j.gastro.2018.09.051
847 Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., . . . Glass, C. K. (2010).
848 Simple combinations of lineage-determining transcription factors prime cis-regulatory
849 elements required for macrophage and B cell identities. Mol Cell, 38(4), 576-589.
850 doi:10.1016/j.molcel.2010.05.004
851 Hoshida, Y., Nijman, S. M., Kobayashi, M., Chan, J. A., Brunet, J. P., Chiang, D. Y., . . . Golub,
852 T. R. (2009). Integrative transcriptome analysis reveals common molecular subclasses of
853 human hepatocellular carcinoma. Cancer Res, 69(18), 7385-7392. doi:10.1158/0008-
854 5472.CAN-09-1089
855 Ji, J., Dai, X., Yeung, S. J., & He, X. (2019). The role of long non-coding RNA GAS5 in cancers.
856 Cancer Manag Res, 11, 2729-2737. doi:10.2147/CMAR.S189052
857 Kedersha, N. L., Gupta, M., Li, W., Miller, I., & Anderson, P. (1999). RNA-binding proteins
858 TIA-1 and TIAR link the phosphorylation of eIF-2 alpha to the assembly of mammalian
859 stress granules. J Cell Biol, 147(7), 1431-1442. doi:10.1083/jcb.147.7.1431
860 Kim, H. S., Headey, S. J., Yoga, Y. M., Scanlon, M. J., Gorospe, M., Wilce, M. C., & Wilce, J.
861 A. (2013). Distinct binding properties of TIAR RRMs and linker region. RNA Biol, 10(4),
862 579-589. doi:10.4161/rna.24341
863 Kim, H. S., Wilce, M. C., Yoga, Y. M., Pendini, N. R., Gunzburg, M. J., Cowieson, N. P., . . .
864 Wilce, J. A. (2011). Different modes of interaction by TIAR and HuR with target RNA
865 and DNA. Nucleic Acids Res, 39(3), 1117-1130. doi:10.1093/nar/gkq837

Kong, L., Zhang, Y., Ye, Z. Q., Liu, X. Q., Zhao, S. Q., Wei, L., & Gao, G. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res, 35(Web Server issue), W345-349. doi:10.1093/nar/gkm391

Lai, M. C., Yang, Z., Zhou, L., Zhu, Q. Q., Xie, H. Y., Zhang, F., . . . Zheng, S. S. (2012). Long non-coding RNA MALAT-1 overexpression predicts tumor recurrence of hepatocellular carcinoma after liver transplantation. Med Oncol, 29(3), 1810-1816. doi:10.1007/s12032-011-0004-z

Lee, S., Kopp, F., Chang, T. C., Sataluri, A., Chen, B., Sivakumar, S., . . . Mendell, J. T. (2016). Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins. Cell, 164(1-2), 69-80. doi:10.1016/j.cell.2015.12.017

Li, C., Chen, J., Zhang, K., Feng, B., Wang, R., & Chen, L. (2015). Progress and Prospects of Long Noncoding RNAs (lncRNAs) in Hepatocellular Carcinoma. Cell Physiol Biochem, 36(2), 423-434. doi:10.1159/000430109

Liao, B., Hu, Y., & Brewer, G. (2007). Competitive binding of AUF1 and TIAR to MYC mRNA controls its translation. Nat Struct Mol Biol, 14(6), 511-518. doi:10.1038/nsmb1249

Lin, M. F., Jungreis, I., & Kellis, M. (2011). PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. Bioinformatics, 27(13), i275-282. doi:10.1093/bioinformatics/btr209

Lin, R., Maeda, S., Liu, C., Karin, M., & Edgington, T. S. (2007). A large noncoding RNA is a marker for murine hepatocellular carcinomas and a spectrum of human carcinomas. Oncogene, 26(6), 851-858. doi:10.1038/sj.onc.1209846

Liu, L., Yue, H., Liu, Q., Yuan, J., Li, J., Wei, G., . . . Chen, R. (2016). LncRNA MT1JP functions as a tumor suppressor by interacting with TIAR to modulate the p53 pathway. Oncotarget, 7(13), 15787-15800. doi:10.18632/oncotarget.7487

Llovet, J. M., Montal, R., Sia, D., & Finn, R. S. (2018). Molecular therapies and precision medicine for hepatocellular carcinoma. Nat Rev Clin Oncol, 15(10), 599-616. doi:10.1038/s41571-018-0073-4

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol, 15(12), 550. doi:10.1186/s13059-014-0550-8

Ma, W., & Mayr, C. (2018). A Membraneless Organelle Associated with the Endoplasmic Reticulum Enables 3'UTR-Mediated Protein-Protein Interactions. Cell, 175(6), 1492-1506 e1419. doi:10.1016/j.cell.2018.10.007

Mazan-Mamczarz, K., Lal, A., Martindale, J. L., Kawai, T., & Gorospe, M. (2006). Translational repression by RNA-binding protein TIAR. Mol Cell Biol, 26(7), 2716-2727. doi:10.1128/MCB.26.7.2716-2727.2006

McHugh, C. A., Chen, C. K., Chow, A., Surka, C. F., Tran, C., McDonel, P., . . . Guttman, M. (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature, 521(7551), 232-236. doi:10.1038/nature14443

Meyer, C., Garzia, A., Mazzola, M., Gerstberger, S., Molina, H., & Tuschl, T. (2018). The TIA1 RNA-Binding Protein Family Regulates EIF2AK2-Mediated Stress Response and Cell Cycle Progression. Mol Cell, 69(4), 622-635 e626. doi:10.1016/j.molcel.2018.01.011

Peng, S. Y., Lai, P. L., & Hsu, H. C. (1993). Amplification of the c-myc gene in human hepatocellular carcinoma: biologic significance. J Formos Med Assoc, 92(10), 866-870.

Perkins, D. N., Pappin, D. J., Creasy, D. M., & Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis, 20(18), 3551-3567. doi:10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2

914 Philip, P. A., Mahoney, M. R., Allmer, C., Thomas, J., Pitot, H. C., Kim, G., . . . Erlichman, C.
915 (2005). Phase II study of Erlotinib (OSI-774) in patients with advanced hepatocellular
916 cancer. J Clin Oncol, 23(27), 6657-6663. doi:10.1200/JCO.2005.14.696
917 Piecyk, M., Wax, S., Beck, A. R., Kedersha, N., Gupta, M., Maritim, B., . . . Anderson, P. (2000).
918 TIA-1 is a translational silencer that selectively regulates the expression of TNF-alpha.
919 EMBO J, 19(15), 4154-4163. doi:10.1093/emboj/19.15.4154
920 Quintela-Fandino, M., Le Tourneau, C., Duran, I., Chen, E. X., Wang, L., Tsao, M., . . . Siu, L. L.
921 (2010). Phase I combination of sorafenib and erlotinib therapy in solid tumors: safety,
922 pharmacokinetic, and pharmacodynamic evaluation from an expansion cohort. Mol
923 Cancer Ther, 9(3), 751-760. doi:10.1158/1535-7163.MCT-09-0868
924 Rimassa, L., & Santoro, A. (2009). Sorafenib therapy in advanced hepatocellular carcinoma: the
925 SHARP trial. Expert Rev Anticancer Ther, 9(6), 739-745. doi:10.1586/era.09.41
926 Rinn, J. L., & Chang, H. Y. (2012). Genome regulation by long noncoding RNAs. Annu Rev
927 Biochem, 81, 145-166. doi:10.1146/annurev-biochem-051410-092902
928 Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., . . . Pappin, D.
929 J. (2004). Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-
930 reactive isobaric tagging reagents. Mol Cell Proteomics, 3(12), 1154-1169.
931 doi:10.1074/mcp.M400129-MCP200
932 Shu, L., Yan, W., & Chen, X. (2006). RNPC1, an RNA-binding protein and a target of the p53
933 family, is required for maintaining the stability of the basal and stress-induced p21
934 transcript. Genes Dev, 20(21), 2961-2972. doi:10.1101/gad.1463306
935 Siegel, R., Ma, J., Zou, Z., & Jemal, A. (2014). Cancer statistics, 2014. CA Cancer J Clin, 64(1),
936 9-29. doi:10.3322/caac.21208

Slaymaker, I. M., Gao, L., Zetsche, B., Scott, D. A., Yan, W. X., & Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. Science, 351(6268), 84-88. doi:10.1126/science.aad5227

Soudyab, M., Iranpour, M., & Ghafouri-Fard, S. (2016). The Role of Long Non-Coding RNAs in Breast Cancer. Arch Iran Med, 19(7), 508-517. doi:0161907/AIM.0011

Tabula Muris, C., Overall, c., Logistical, c., Organ, c., processing, Library, p., . . . Principal, i. (2018). Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature, 562(7727), 367-372. doi:10.1038/s41586-018-0590-4

Tani, H., Mizutani, R., Salam, K. A., Tano, K., Ijiri, K., Wakamatsu, A., . . . Akimitsu, N. (2012). Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. Genome Res, 22(5), 947-956. doi:10.1101/gr.130559.111

Tsai, M. C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J. K., Lan, F., . . . Chang, H. Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. Science, 329(5992), 689-693. doi:10.1126/science.1192002

Uhlen, M., Zhang, C., Lee, S., Sjostedt, E., Fagerberg, L., Bidkhori, G., . . . Ponten, F. (2017). A pathology atlas of the human cancer transcriptome. Science, 357(6352). doi:10.1126/science.aan2507

Villanueva, A. (2019). Hepatocellular Carcinoma. N Engl J Med, 380(15), 1450-1462. doi:10.1056/NEJMra1713263

Wang, J., Liu, X., Wu, H., Ni, P., Gu, Z., Qiao, Y., . . . Fan, Q. (2010). CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer. Nucleic Acids Res, 38(16), 5366-5383. doi:10.1093/nar/gkq285

959 Wang, L., Park, H. J., Dasari, S., Wang, S., Kocher, J. P., & Li, W. (2013). CPAT: Coding-
960 Potential Assessment Tool using an alignment-free logistic regression model. Nucleic
961 Acids Res, 41(6), e74. doi:10.1093/nar/gkt006
962 Yamada, N., Hasegawa, Y., Yue, M., Hamada, T., Nakagawa, S., & Ogawa, Y. (2015). Xist
963 Exon 7 Contributes to the Stable Localization of Xist RNA on the Inactive X-
964 Chromosome. PLoS Genet, 11(8), e1005430. doi:10.1371/journal.pgen.1005430
965 Yu, F. J., Zheng, J. J., Dong, P. H., & Fan, X. M. (2015). Long non-coding RNAs and
966 hepatocellular carcinoma. Mol Clin Oncol, 3(1), 13-17. doi:10.3892/mco.2014.429
967 Yuan, S. X., Wang, J., Yang, F., Tao, Q. F., Zhang, J., Wang, L. L., . . . Zhou, W. P. (2016).
968 Long noncoding RNA DANCR increases stemness features of hepatocellular carcinoma
969 by derepression of CTNNB1. Hepatology, 63(2), 499-511. doi:10.1002/hep.27893
970 Zhao, R., Nakamura, T., Fu, Y., Lazar, Z., & Spector, D. L. (2011). Gene bookmarking
971 accelerates the kinetics of post-mitotic transcriptional re-activation. Nat Cell Biol, 13(11),
972 1295-1304. doi:10.1038/ncb2341
973 Zheng, K., Cubero, F. J., & Nevzorova, Y. A. (2017). c-MYC-Making Liver Sick: Role of c-
974 MYC in Hepatic Cell Function, Homeostasis and Disease. Genes (Basel), 8(4).
975 doi:10.3390/genes8040123

976

977

Figure 1. LncRNA screen to identify transcripts enriched in ESCs and dysregulated in cancer

A. PCA plot of 10 RNA-seq libraries from mouse derived ESCs, and two from cell lines. Differentiation from ESCs to NPCs created the largest difference in variance, while there was minimal difference between isolated clones vs. cell lines.
B. Workflow of the filtering process performed to obtain ESC enriched lncRNAs that are also dysregulated in cancer. Red indicates analysis performed in mouse and blue indicates human.
C. LncRNA candidate expression across ENCODE tissue datasets show that lncRNAs are mostly not pan-expressed, but are rather tissue specific. Counts are scaled per row.
D. *LINC00862* is upregulated in both human cirrhotic liver and HCC tumor samples when compared to control patient liver tissue samples. **p < 0.01; ***p < 0.005; Student's t-test.
E. *LINC00862* is upregulated in various human HCC cell lines.

Figure 2. *PHAROH* lncRNA is highly expressed in ESCs, embryonic liver, models of hepatocarcinogenesis, and HCC cell lines

A. 5' 3' RACE reveals two isoforms for *PHAROH*, which have exons 3 and 4 in common. *PHAROH* is an intergenic lncRNA where the nearest upstream gene is *Zfp218* (51 kb away), and downstream is *Nr5a2* (151 kb away). RNA-seq tracks of ESC (red) and Hepa1-6 (blue) cells show cell-type specific isoform expression of *PHAROH*.
B. *PHAROH* is highly expressed in embryonic liver in E14 and E18 mice, but not adult liver (**p < 0.01; ***p < 0.005; Student's t-test).
C. A DEN model of hepatocarcinogenesis shows high upregulation of *PHAROH* in the liver and HCC tumor nodules (gray bar) in DEN treated mice (**p < 0.01; ***p < 0.005; Student's t-test.).
D. *PHAROH* is upregulated in HCC cell lines (Hepa1-6, and Hepa1c1c7) compared to normal mouse hepatocytes (AML12) (***p < 0.005; Student's t-test).

E. Single molecule RNA-FISH of *PHAROH* in ESCs shows nuclear localization and an average of 3-5 foci per cell. In Hepa1-6 cells, *PHAROH* shows 25 foci per cell, distributed evenly between the nucleus and cytoplasm (n=75 cells for each sample). *Ppib* is used as a housekeeping protein coding gene control.
F. Quantitation of panel *PHAROH* foci in panel E in HepA1-6 cells
G. Cellular fractionation of Hepa1-6 cells shows equal distribution of *PHAROH* in the cytoplasm and nucleus, where it also binds to chromatin. *Gapdh* is predominantly cytoplasmic, and *MALAT1* is bound to chromatin.

Figure 3. Depletion of *PHAROH* results in a proliferation defect

A. Four isolated clones all have a comparable deletion of 788 bp. The wildtype band is ~1.8 kb.
B. qRT-PCR of *PHAROH* knockout clones show a >80% reduction in *PHAROH* levels (***p < 0.005; Student's t-test).
C. Aggregated doubling time of clones shows knockout of *PHAROH* increases doubling time from 14.8h to 18.6h. Addition of *PHAROH* back into knockouts rescues this defect (***p < 0.005; Student's t-test).
D. Manual cell counting shows proliferation defect in *PHAROH* knockout cells that is rescued by ectopic expression of *PHAROH* (*p < 0.05; Student's t-test).
E. Migration distance for *PHAROH* knockout clones is decreased by 50% (**p < 0.01; Student's t-test).
F. 50% Knockdown of *PHAROH* can be achieved using both ASO7 and ASO15 at 24h (***p < 0.005; Student's t-test).
G. Colony formation assay of Hepa1-6 cells that are treated with scrambled or *PHAROH* targeting ASOs. After seeding 200 cells and two weeks of growth, a 50% reduction in relative colony number is observed (**p < 0.01; Student's t-test).

Figure 4. Gene expression analysis of *PHAROH* knockout cells reveals a link to MYC

A. GO term analysis of differentially expressed genes shows enrichment of cell

proliferation and migration genes

B. qRT-PCR of Myc mRNA levels indicate that Myc transcript does not appreciably change when *PHAROH* is knocked out.

C. Western blot analysis of MYC protein shows downregulation of protein levels in *PHAROH* knockout cells. β-ACTIN is used as a loading control.

D. qRT-PCR of genes downstream of Myc shows a statistically significant decrease in expression (*$p < 0.05$; **$p < 0.01$; Student's t-test).

Figure 5. RAP-MS identifies TIAR as a major interactor of *PHAROH*

A. Five different biotinylated oligos antisense to *PHAROH* were screened for pulldown efficiency. Oligos 2-5 can pull down *PHAROH* at ~80% efficiency or greater

B. *PHAROH* can be eluted at a specific temperature. Maximum elution is reached at 40° C.

C. iTRAQ results using two different oligos targeting *PHAROH* compared to PPIB reveal nucleolysin TIAR as the top hit.

D. TIAR is pulled down by *PHAROH* oligos, and is specifically eluted at 40° C, but not by PPIB oligos.

E. TIAR can be pulled down using additional oligos and in two different cell lines. RNase A treatment of the protein lysate diminishes TIAR binding to *PHAROH*, indicating that the interaction is RNA-dependent.

F. Immunoprecipitation of TIAR enriches for *PHAROH* transcript, when compared to IgG and PPIB control (***$p < 0.005$; Student's t-test).

Figure 6. TIAR binds to the 5' end of PHAROH

A. Sequence analysis of *PHAROH* with published TIAR binding motifs shows a preference for the 5' end of *PHAROH*.

B. Schematic of the conserved hairpin of *PHAROH* that contains four potential TIAR binding sites indicated in the red boxes. Mutations created within the *PHAROH* hairpin are indicated in red asterisks.

C. RNA EMSA of the 71-nt *PHAROH* hairpin with human recombinant TIAR shows three sequential shifts as TIAR concentration increases.

D. Densitometry analysis of the free unbound probe estimates the dissociation constant of TIAR as ~2 nM.

E. TIAR/PHAROH binding is specific as a supershift is created when adding antibody against TIAR, and the interaction can be competed out using 20x unlabeled RNA. RNA EMSA of the mutant hairpins reveals decreasing affinity for TIAR. Mutants were made in a cumulative 5' to 3' fashion. M1 shows high signal of single and double occupancy forms, and m2 has reduced signal overall. When all four sites are mutated, binding is nearly abolished.

Figure 7. Loss of *PHAROH* releases TIAR, which inhibits Myc translation

A. RNA EMSA of the 53-nt Myc 3' UTR fragment shows that TIAR has three potential binding sites, but prefers a single binding event (note arrows)

B. Knockdown of *PHAROH* reduces MYC protein levels, but not TIAR levels, even though MYC is expressed 3-fold higher than *PHAROH*.

C. Wildtype *PHAROH* hairpin is able to compete out the MYC-TIAR interaction, but the mutated hairpin is not as effective in competing with the Myc-TIAR interaction.

D. Luciferase activity is increased with the addition of *PHAROH* but not with m_4*PHAROH* (**p < 0.01; Student's t-test).

E. Overexpression of *PHAROH* increases MYC protein expression, but overexpression of m_4*PHAROH* does not change MYC levels appreciably.

Figure 1-figure supplement 1.

A. LncRNA screen identifies candidates with varying levels of expression in ESCs.
B. *LINC00862* is altered in 13% of all HCC patient cases according to TCGA data.
C. *Gm19705* gene locus on chromosome 1 shows that the order of the genes is conserved between mouse and human, but the order is reversed, suggesting a reversal event occurrence.
D. Weighted gene correlation network analysis of *Gm19705* places it in a module with cell cycle genes and proliferation genes, such as *Brca1/2*, and *Mki67*.
E. GO term analysis of the module containing *Gm19705* shows enrichment of genes related to cell cycle, mitosis, and DNA replication.
F. Re-analysis of single cell data of adult liver (Tabula Muris et al., 2018) reveals expression of *Gm19705* is highly enriched in hepatocytes, but only a subset of the cells.

Figure 2-figure supplement 2.

A. CPC and CPAT coding potential analysis for *PHAROH* and *Gapdh*.
B. PhyloCSF tracks showing low coding potential for the *PHAROH* locus
C. *PHAROH* is expressed in fetal liver, but not in the blood.
D. Sham hepatectomy (SH) or partial hepatectomy (PH) of the liver, a model of liver regeneration, shows upregulation of *PHAROH* during time points of concerted cell division.
   *p < 0.05; **p < 0.01; ***p < 0.005; Student's t-test.
E. HepA1-6 cells were stained with Hoechst 33258 and sorted according to their cell cycle phase. qRT-PCR analysis shows that *PHAROH* does not cycle with the cell cycle, unlike *Ccnb1* and *Ccne1*.
F. qRT-PCR of each annotated exon spanning the current Gencode M20 annotation. Exons 1-4, which are numbered similarly as Figure 2A, are confirmed RACE exons. Isoform with exons 1, 3, and 4 is ESC specific, and the isoform with exons 2-4 is HCC specific. Exons A, B, C, D, and E are currently annotated exons, but not identifiable via RACE.
G. *XIST*, *FIRRE*, and *NEAT1* serve as additional controls for the cellular fractionation.

H. Calculated RNA half-life based upon α-amanitin treated cells. *PHAROH* has a half-life of 10.8h, longer than that of *XIST* and *MALAT1*.

Figure 3-figure supplement 3.

A. FACS for double GFP+/mCherry+ cells shows an 85.1% nucleofection efficiency for both plasmids.
B. Knockdown of *PHAROH* using nucleofection of 2 μM ASO is effective over 96h.
C. MTS assay for proliferation 96h after nucleofection. MTS absorbance is reduced by 50% in ASO treated samples targeting *PHAROH* and *Eg5*.
D. Reduction of colony formation number is dose dependent.

Figure 4-figure supplement 4.

A. Principal component analysis of two sgRenilla negative control clones and two *PHAROH* knockout clones. Deletion of *PHAROH* is well separated by PC1.
B. Euclidean distance plot indicating that the negative control clones and *PHAROH* knockout clones cluster independently.
C. Motif analysis of promoter region of differentially expressed genes. MYC motif is enriched 1.24 fold over background sequences.
D. Immunofluorescence of MYC in *PHAROH* knockout clones shows absence of MYC signal in a majority of cells. Scale bar = 50 μm.
E. Quantification of western blot in Figure 4C.

Figure 5-figure supplement 5.

A. The amount of *PHAROH* RNA remaining on the beads after thermal elution is inverse to that of the eluate.
B. Off-target pulldown of *Ppib* using *PHAROH* oligos is low.
C. An oligo designed against *Ppib* can pull the RNA down at ~65% efficiency, and does not pull down *PHAROH* or 18S.
D. *Ppib* can also be eluted via a temperature gradient, and is optimally released at 40° C.

Figure 6-figure supplement 6.

A. Mapping the top seven binding sites to predicted structures (top three shown here), reveals a conserved hairpin on the majority of predicted structures.
B. Profile analysis of the RNA EMSA gel in Fig. 6C, showing the shift in intensity.
C. Binding of TIAR to m2 and m3 are similar, possibly due to the mutation of a weaker binding site does not greatly impact overall binding.

Figure 7-figure supplement 7.

A. Of the two TIAR binding sites on MYC's 3' UTR, only one maps to the mouse genome.
B. Potential TIAR binding sites on the mouse Myc 3' UTR highlighted in red.
C. Knockdown of *PHAROH* does not change Myc mRNA levels, suggesting that *PHAROH* acts at a post-transcriptional level.
D. Addition of *PHAROH* to a luciferase construct with a Myc 3' UTR increases luciferase activity in a dose dependent manner.
E. MYC RNA levels do not change when *PHAROH* or TIAR are overexpressed.
F. IF microscopy of TIAR showing predominantly nuclear localization. Scale bar = 25 μm
G. Kaplan-Meier survival plot of patients with low and high TIAR expression.

Table 1

Gene Name	Sequence Homology	Synteny	Human Homologue
Platr15	-	+	LOC284798
4930444M15Rik	64.4% of bases, 99.9% of span	+	In TUSC8 region
5430416N02Rik	16.6% of bases, 100.0% of span	+	Thap9-AS1
Platr6	45.2% of bases, 85.5% of span	+	LINC01010
6720427I07Rik	94.3% of bases, 100.0% of span	+	LINC02603
B830012L14Rik	57.4% of bases, 83.8% of span	+	Meg8 (GM26945)
C330004P14Rik	-	+	LINC01625
Gm38509	22.9% of bases, 84.4% of span	+	LINC01206
A330094K24Rik	54.7% of bases, 100.0% of span	+	C18orf25 (PCG)
Bvht	53.2% of bases, 100.0% of span	+	Carmn
Dancr	48.2% of bases, 49.0% of span	+	Dancr
2900041M22Rik	50.2% of bases, 60.5% of span	+	LINC01973
Dleu2	72.8% of bases, 100.0% of span	+	Dleu2
E130202H07Rik	61.7% of bases, 65.2% of span		Tusc8
Epb41l4aos	69.0% of bases, 100.0% of span	+	Epb41l4a-AS1
Firre	7.0% of bases, 14.5% of span	+	Firre
Gm20939	-	+	LINC00470
Gas5	71.3% of bases, 97.7% of span	+	Gas5
Gm12688	92.6% of bases, 100.0% of span	+	FOXD3-AS1
Gm47599	21.6% of bases, 85.0% of span	+	Socs2-AS1
Gm19705	27.6% of bases, 47.8% of span	+	LINC00862
Gm20703	79.2% of bases, 100.0% of span	+	GAPLINC
Gm26763	3.6% of bases, 3.8% of span	+	Smarca5-AS1
Gm26945	65.4% of bases, 67.8% of span	+	Meg8
AC129328.1	-	+	LINC01340,
Gm28373	44.6% of bases, 83.5% of span	+	Itpk1-AS1
Gm31693	12.7% of bases, 24.9% of span	+	LINC00578
Mir124a-1hg	91.7% of bases, 100.0% of span	+	LINC00599
Mir142hg	74.5% of bases, 100.0% of span	+	TSPOAP1-AS1
Mir17hg	74.7% of bases, 100.0% of span	+	Mir17Hg
Neat1	37.5% of bases, 100.0% of span	+	NEAT1
Platr12	16.2% of bases, 33.7% of span	+	GPR1-AS
Rbakdn	96.4% of bases, 99.1% of span	+	Rbakdn
Snhg1	73.3% of bases, 89.2% of span	+	Snhg1
Snhg14	4.5% of bases, 5.4% of span	+	Snhg14
D5Ertd605e	-	+	Pan3-AS1
Snhg18	83.3% of bases, 100.0% of span	+	Snhg18
Snhg5	67.8% of bases, 81.6% of span	+	Snhg5
Sptbn5	78.8% of bases, 100.0% of span	+	Sptbn5
Xist	70.1% of bases, 100.0% of span	+	Xist

Table 2

Protein Hit	Ratio
Tial1	2.15559
Hnrnpab	1.80692
Rbm3	1.77037
Hnrnpd	1.62883
Hnrnpa1	1.6283
Ptbp2	1.57804
Hnrnpa3	1.53035
Caprin1	1.50299
Lmna	1.37542
Fubp3	1.34941
Banf1	1.34137
Hnrnpa2b1	1.33969
H2afj	1.3213
Lima1	1.20909
Nolc1	1.20733
Abcb5	1.19592
Nup62	1.18297
Elavl1	1.09477
Ssbp1	1.08439
Hist1h2bc	1.07366
Itgax	1.00222
Rbm8a	0.98396
Dhx9	0.95827
Smu1	0.94938
Cnbp	0.9225
Nup93	0.82199
Lsm3	0.79027
Xrcc5	0.78242
Med25	0.76892
Actc1	0.76507
Khsrp	0.75921
Actb	0.75109
Nipsnap1	0.75014
Pnn	0.74713
Hba-a1	0.74299
Snrpe	0.74052
Nol11	0.73772
Erh	0.73354
Psmb1	0.72391
Efhd2	0.71468

1183

1 Key Resources Table

Reagent Type	Designation	Source or reference	Identifiers	Additional Information
Strain, strain background (C57BL/6J) *Mus musculus* Female	C57BL/6J	The Jackson Laboratory	Stock No: 000664 RRID:IMSR_JAX: 000664	
Gene (*Homo sapiens*)	Tial1 (NM_009383) Mouse Tagged ORF Clone	Origene	Cat# MG226372	
Gene (*Mus musculus*)	Myc	GenBank	NC_000081.7	
Recombinant Protein (*Homo sapiens*)	Recombinant Human TIAL1 Protein	Novus Biologicals	Cat# NBP2-51914-0.1mg	
Cell line (*Mus musculus*)	AB2.2 (ESCs)	Bergmann et al., 2015		Cell line maintained in D. L. Spector Lab
Cell line (*Mus musculus*)	NPC	Bergmann et al., 2015		Cell line maintained in D. L. Spector Lab
Cell line (*Mus musculus*)	Hepa1-6	ATCC	Cat# CRL-1830	Cell line maintained in D. L. Spector Lab
Cell line (*Mus musculus*)	Hepa1c1c7	ATCC	Cat# CRL-2026	Cell line maintained in D. L. Spector Lab
Cell line (*Mus musculus*)	AML12	ATCC	Cat# CRL-2254	Cell line maintained in D. L. Spector Lab
Cell line (*Mus musculus*)	MEF	MTI-Global Stem	Cat# GSC-6601G	Irradiated feeder MEFs
Cell line (*Homo sapiens*)	SNU-182	ATCC	Cat# CRL-2235	Cell line maintained in D. L. Spector Lab
Cell line (*Homo sapiens*)	Huh1	N/A		Generous gift from Scott Lowe (MSKCC)
Cell line (*Homo sapiens*)	Huh7	N/A		Generous gift from Scott Lowe (MSKCC)
Cell line (*Homo sapiens*)	JHH2	N/A		RNA gifted from Scott Lowe (MSKCC)
Cell line (*Homo sapiens*)	SNU-387	ATCC	Cat# CRL-2237	Generous gift from Scott Lowe (MSKCC)
Cell line (*Homo*	Hep3B	ATCC	Cat# HB-8064	Generous gift

sapiens)				from Scott Lowe (MSKCC)
Cell line (*Homo sapiens*)	Alex	ATCC	Cat# CRL-8024	RNA gifted from Scott Lowe (MSKCC)
Cell line (*Homo sapiens*)	HepG2	ATCC	Cat# HB-8065	Generous gift from Scott Lowe (MSKCC)
Cell line (*Homo sapiens*)	Li7	N/A		RNA gifted from Scott Lowe (MSKCC)
Cell line (*Homo sapiens*)	THLE-2	ATCC	Cat# CRL-2706	Cell line maintained in D. L. Spector Lab
Antibody	c-Myc, (rabbit monoclonal)	Cell SIgnaling	Cat# 5605 RRID:AB_1903938	(IB: 1:1,000)
Antibody	TIAR, (rabbit monoclonal)	Cell Signaling	Cat# 8509 RRID:AB_10839263	(IB: 1:1,000) (IF: 1:2,000) (IP: 1:100)
Antibody	β-Actin, (mouse monoclonal)	Cell Signaling	Cat# 3700 RRID:AB_2242334	(IB: 1:10,000)
Antibody	IRDye 800CW (Goat anti-Rabbit IgG)	LI-COR Biosciences	Cat# 925-32211 RRID:AB_2651127	(IB: 1:10,000)
Antibody	IRDye 680RD (Goat anti-Mouse IgG)	LI-COR Biosciences	Cat# 925-68070 RRID:AB_2651128	(IB: 1:10,000)
Antibody	Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody Alexa Fluor 488	Thermo Fisher	Cat# A-11008 RRID:AB_143165	(IF: 1:1,000)
Antibody	Rabbit IgG Isotype Control	Thermo Fisher	Cat# 10500C RRID:AB_2532981	
Recombinant DNA reagent	eSpCas9-1.1	Addgene	RRID:Addgene_71814	Backbone for constructing GFP and mCherry variants
Recombinant DNA reagent	eSpCas9-1.1-GFP (plasmid)	This study	N/A	
Recombinant DNA reagent	eSpCas9-1.1-mCherry (plasmid)	This study	N/A	
Recombinant DNA reagent	pmirGLO	Promega	Cat# E1330	Dual-Luciferase miRNA Target Expression

				Vector
Recombinant DNA reagent	pCMV6-A-Puro	Origene	Cat# PS100025	pCMV6 backbone
Transfected construct (*Mus musculus*)	sgPHAROH_F-eSpCas9-1.1-GFP (plasmid)	This study	N/A	Upstream PHAROH sgRNA
Transfected construct (*Mus musculus*)	sgPHAROH_R-eSpCas9-1.1-mCherry (plasmid)	This study	N/A	Downstream PHAROH sgRNA
Transfected construct (*Mus musculus*)	sgRenilla-eSpCas9-1.1-GFP (plasmid)	Chang et al., 2020	N/A	Negative control sgRNA
Transfected construct (*Mus musculus*)	pmirGLO-MYC (plasmid)	This study	N/A	Contruct for luciferase assay readout
Transfected construct (*Mus musculus*)	pCMV6-pharoh (plasmid)	This study	N/A	Contruct for rescue and luciferase assay readout
Transfected construct (*Mus musculus*)	pCMV6-m4pharoh (plasmid)	This study	N/A	Contruct for luciferase assay readout
Transfected construct (*Mus musculus*)	pCMV6-GFP (plasmid)	Chang et al., 2020	N/A	Contruct for luciferase assay readout
Sequence-based reagent	ASO 7	This study	*PHAROH* Gapmer ASO	CGTGTCATCTTC TTGGCCCC
Sequence-based reagent	ASO 15	This study	*PHAROH* Gapmer ASO	TCGTGTCATCTT CTTGGCCC
Sequence-based reagent	ASO 14	This study	*PHAROH* cEt ASO	GTTACAGGACG CATGT
Sequence-based reagent	ASO 18	This study	*PHAROH* cEt ASO	CACATAGTTATT CCCG
Sequence-based reagent	Forward	This study	*PHAROH* genomic PCR	TGCTTAGCACGT CCTCAGTGC
Sequence-based reagent	Reverse	This study	*PHAROH* genomic PCR	AGTTCCCCAGCA ACCCTGTT
Sequence-based reagent	Upstream	This study	*PHAROH* sgRNA	GCAGGTAGTGT GGTAACTCC
Sequence-based reagent	Downstream	This study	*PHAROH* sgRNA	CGGGTCCTCCCA GCGCACAC
Sequence-based reagent	Exon 4 Fwd	This study	*PHAROH* qRT-PCR	GGGGCCAAGAA GATGACACG
Sequence-based reagent	Exon 4 Ref	This study	*PHAROH* qRT-PCR	GGACGCATGTG GAGGTCAGA
Sequence-based reagent	Exon A Fwd	This study	*PHAROH* qRT-PCR	TGCCTCACAAGG GACAACACTC
Sequence-based	Exon A Rev	This study	*PHAROH* qRT-	GAATTTGCTCAG

reagent			PCR	GGGCTCCA
Sequence-based reagent	Exon B Fwd	This study	*PHAROH* qRT-PCR	GGACTTGAACTG GCACTGTTGC
Sequence-based reagent	Exon B Rev	This study	*PHAROH* qRT-PCR	CAGAAGGACCA TCATCACGA
Sequence-based reagent	Exon C Fwd	This study	*PHAROH* qRT-PCR	TGAACCCGAGCT TTGCCATT
Sequence-based reagent	Exon C Rev	This study	*PHAROH* qRT-PCR	CGGTGCTCTGCA GGACGTTT
Sequence-based reagent	Exon D Fwd	This study	*PHAROH* qRT-PCR	AGGCTGCCGCC ACACTTAAA
Sequence-based reagent	Exon D Rev	This study	*PHAROH* qRT-PCR	TTCAGCTGCTGG CATTCTTCC
Sequence-based reagent	Exon E Fwd	This study	*PHAROH* qRT-PCR	GGAGAGAACAA GGGCCTTCC
Sequence-based reagent	Exon E Rev	This study	*PHAROH* qRT-PCR	GCCCTGCTGCAT TCTGGGTA
Sequence-based reagent	Exon 1 Fwd	This study	*PHAROH* qRT-PCR	GGTGTGAACCA AGTGCACGTCT
Sequence-based reagent	Exon 1 Rev	This study	*PHAROH* qRT-PCR	GGGATCTGACA CCGCCTTCTT
Sequence-based reagent	Exon 2 Fwd	This study	*PHAROH* qRT-PCR	CTTCTGAGTCTG ACGGGCTGGT
Sequence-based reagent	Exon 2 Rev	This study	*PHAROH* qRT-PCR	TCAGTCCTACCC AAGAAATTTAG GA
Sequence-based reagent	Exon 3 Fwd	This study	*PHAROH* qRT-PCR	TGTGGAAACTCA GAGAGGATGC
Sequence-based reagent	Exon 3 Rev	This study	*PHAROH* qRT-PCR	CTCTGGTGGCTG TGCCTTCAAA
Sequence-based reagent	MycF	This study	Myc qRT-PCR	CAACGTCTTGGA ACGTCAGA
Sequence-based reagent	MycR	This study	Myc qRT-PCR	TCGTCTGCTTGA ATGGACAG
Sequence-based reagent	Outer 1	This study	5' RACE	TTCCTGCGTGAA AGTGTCTG
Sequence-based reagent	Outer 2	This study	5' RACE	TGACCTTCTCAG GAAGTGGAA
Sequence-based reagent	Inner 1	This study	5' RACE	CCTGAGAGGAC GAGGTGACT
Sequence-based reagent	Inner 2	This study	5' RACE	TTTGCAGGTTAG GATCAGAGC
Sequence-based reagent	Outer	This study	3' RACE	CACTTCCATTCCT CCCCATA
Sequence-based reagent	Inner	This study	3' RACE	GGGGACTCAGA CACTCACCA
Sequence-based reagent	PHAROH hairpin	This study	T7 Transcription Primer	TAATACGACTCA CTATA gagaggatgccactg ttttgaactattttgaa

				ggcacagccaccag agctttagggacagg gtattttatc
Sequence-based reagent	Myc 3' UTR	This study	T7 Transcription Primer	TAATACGACTCA CTATAG cttcccatctttttttctt tttccttttaacagatt tgtatttaattgttttt
Sequence-based reagent	m1	This study	T7 Transcription Primer	TAATACGACTCA CTATA gagaggatgccactg tCtCgaactattttga aggcacagccacca gagctttagggacag ggtattttatc
Sequence-based reagent	m2	This study	T7 Transcription Primer	TAATACGACTCA CTATA gagaggatgccactg tCtCgaactaCtCtg aaggcacagccacc agagctttagggaca gggtattttatc
Sequence-based reagent	m3	This study	T7 Transcription Primer	TAATACGACTCA CTATA gagaggatgccactg tCtCgaactaCtCtg aaggcacagccacc agagcCttagggaca gggtattttatc
Sequence-based reagent	m4	This study	T7 Transcription Primer	TAATACGACTCA CTATA gagaggatgccactg tCtCgaactaCtCtg aaggcacagccacc agagcCttagggaca gggtatCCtatc
Sequence-based reagent	PHAROH 1	This study	Biotin antisense pulldown oligo	AGAAATTTAGG AGCCACGCT
Sequence-based reagent	PHAROH 2	This study	Biotin antisense pulldown oligo	GCTGTGCCTTCA AAATAGTT
Sequence-based reagent	PHAROH 3	This study	Biotin antisense pulldown oligo	GCCCCAAGAAA CTCAAGAAT
Sequence-based reagent	PHAROH 4	This study	Biotin antisense pulldown oligo	TTAATTTTCTCCT TTATGCA
Sequence-based reagent	PHAROH 5	This study	Biotin antisense pulldown oligo	ACAACGTGTGG ATGTGTGTT
Sequence-based reagent	PPIB 1	This study	Biotin antisense pulldown oligo	CCTACAGATTCA TCTCCAAT
Sequence-based reagent	PPIB 2	This study	Biotin antisense pulldown oligo	GTTATGAAGAAC TGTGAGCC
Commercial	DNase I,	Life	Cat# 18068	

assay or kit	Amplification Grade	Technologies		
Commercial assay or kit	TaqMan Reverse Transcription Reagents	Thermo Fisher	Cat# 4304134	
Commercial assay or kit	SF Cell Line 4D-Nucleofector X Kit L	Lonza	Cat# V4XC-2024	
Commercial assay or kit	View ISH Cell Assay Kit	Affymetrix	Cat# QVC0001	
Commercial assay or kit	MEGAscript™ T7 Transcription Kit	Thermo Fisher	AM1333	
Commercial assay or kit	Pierce™ RNA 3' End Biotinylation Kit	Thermo Fisher	Cat# 20160	
Commercial assay or kit	LightShift™ Chemiluminescent RNA EMSA Kit	Thermo Fisher	Cat# 20158	
Commercial assay or kit	Pierce BCA Protein Assay Kit	Life Technologies	Cat# 23227	
Commercial assay or kit	CellTiter 96 AQueous One Solution Cell Proliferation Assay	Promega	Cat# G3582	
Commercial assay or kit	SMARTer RACE 5’/3’ Kit	Takara	Cat# 634858	
Commercial assay or kit	Promega Dual-Luciferase® Reporter Assay System	Promega	Cat# E1960	
Commercial assay or kit	DNeasy Blood and Tissue kit	Qiagen	Cat# 69504	
Software, algorithm	Benchling	https://www.benchling.com/	Used for sgRNA design and cloning	
Software, algorithm	CPAT	doi: 10.1093/nar/gkt006		
Software, algorithm	CPC	doi: 10.1093/nar/gkm391		
Software, algorithm	PhyloCSF	doi: 10.1093/bioinformatics/btr209		
Software, algorithm	FastQC	https://www.bioinformatics.babraham.ac.uk/p	RRID:SCR_014583	

		rojects/fastqc/		
Software, algorithm	STAR	doi: 10.1002/0471250953.bi1114s51	RRID:SCR_004463	
Software, algorithm	DESeq2	doi: 10.1186/s13059-014-0550-8	RRID:SCR_015687	
Software, algorithm	Volocity 3D Image Analysis Software	Perkin Elmer	RRID:SCR_002668	
Software, algorithm	SoftWoRx	SoftWoRx Software	RRID:SCR_019157	
Software, algorithm	Sequest HT	doi: 10.1016/1044-0305(94)80016-2		
Software, algorithm	Mascot 2.5	doi: 10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2	RRID:SCR_014322	
Software, algorithm	HOMER Suite	doi: 10.1016/j.molcel.2010.05.004	RRID:SCR_010881	
Software, algorithm	Image Studio Software	LI-COR	RRID:SCR_015795	
Software, algorithm	RNAfold	doi: 10.1093/nar/gkg599	RRID:SCR_008550	
Software, algorithm	mFold	doi: 10.1093/nar/gkg595	RRID:SCR_008543	
Software, algorithm	ImageJ	NIH, Bethesda, MD	RRID:SCR_003070	
Chemical compound, drug	Hoechst dye	Thermo Fisher	Cat# 62249	1 ug/mL
Chemical compound, drug	DAPI	Life Technologies	Cat# D1306	1 ug/mL
Chemical compound, drug	α-Amanitin	Sigma-Aldrich	Cat# A2263	5 ug/mL
Chemical compound, drug	Diethylnitrosamine	Sigma-Aldrich	Cat# 73861	25 mg/kg

2

Figure 1

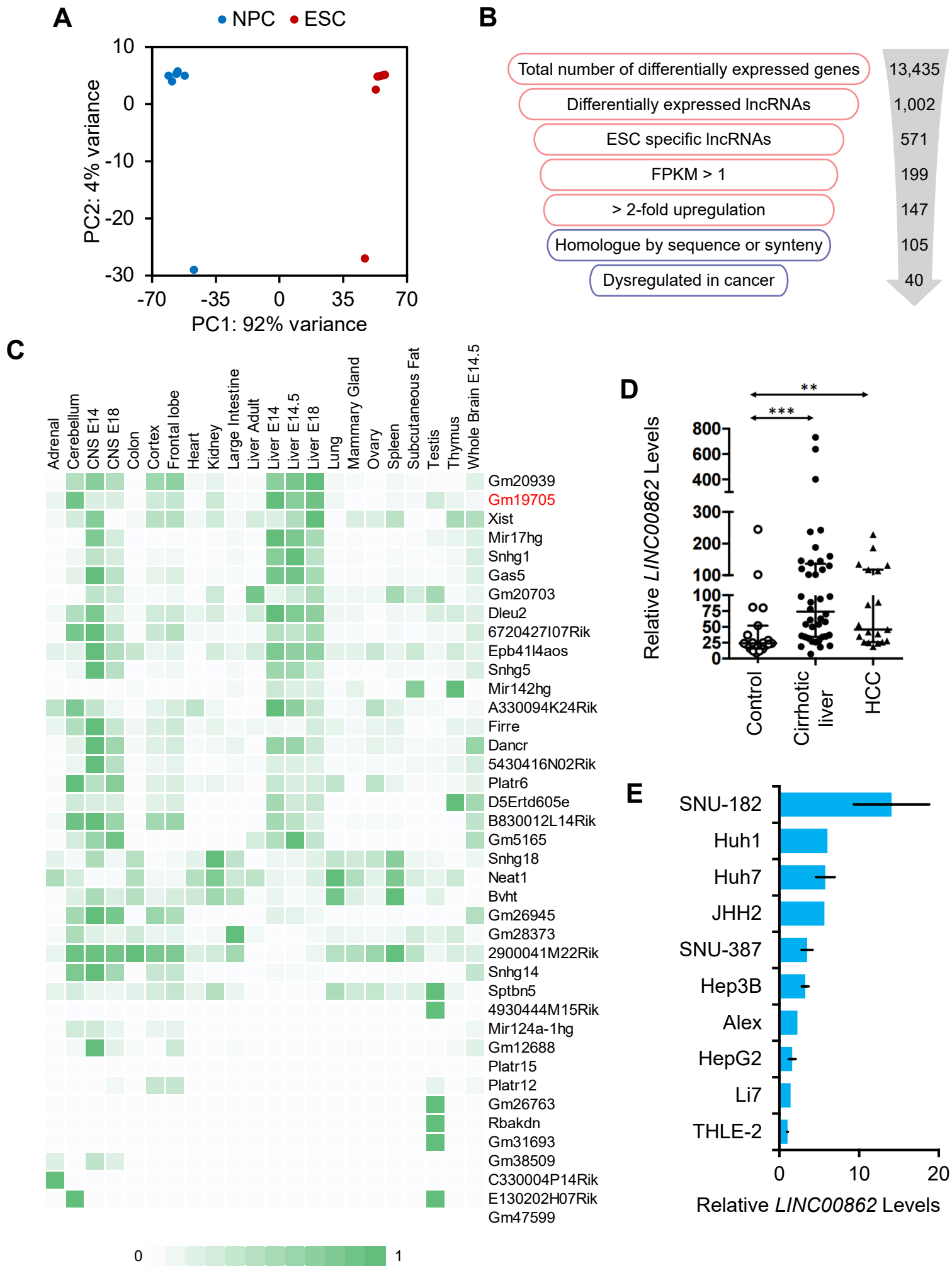
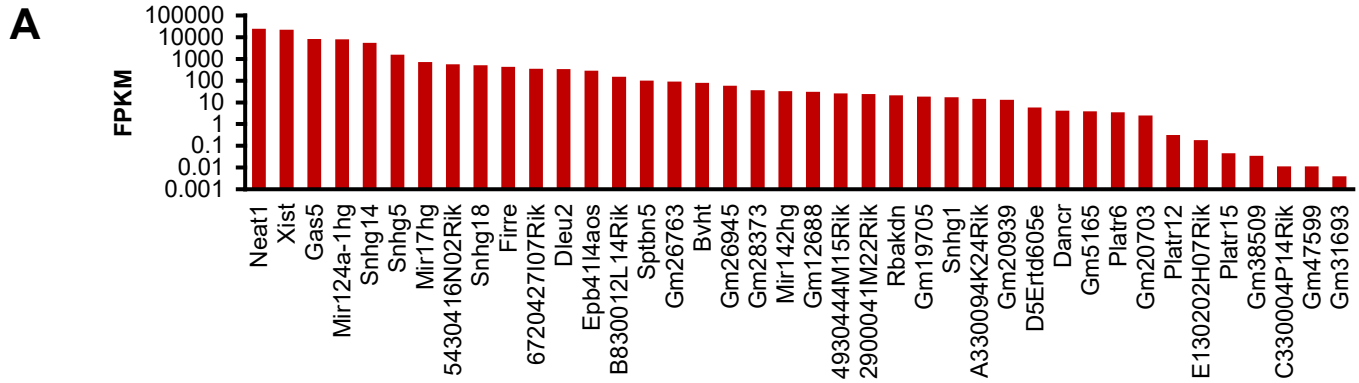
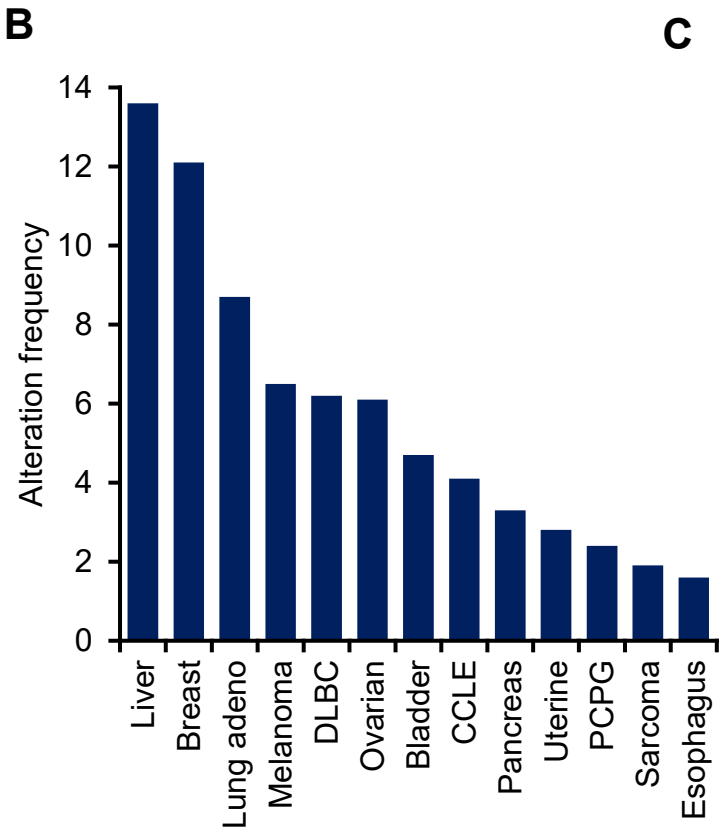
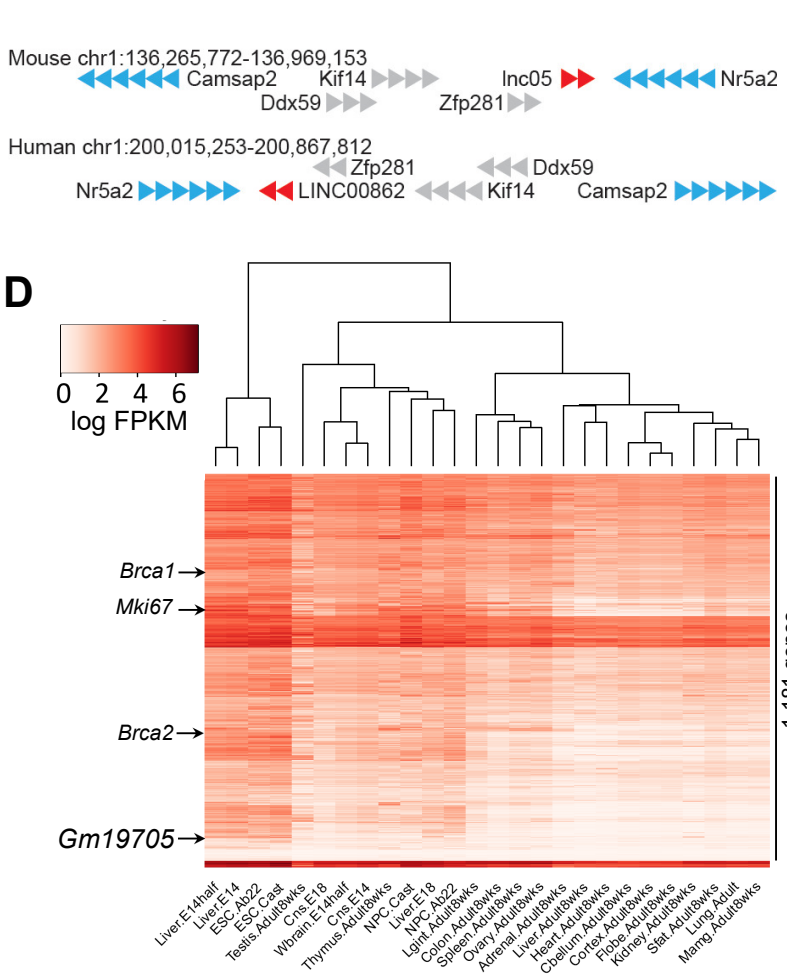

A
NPC
ESC
PC2: 4% variance
PC1: 92% variance
B
Total number of differentially expressed genes
Differentially expressed lncRNAs
ESC specific lncRNAs
FPKM > 1
> 2-fold upregulation
Homologue by sequence or synteny
Dysregulated in cancer
13,435
1,002
571
199
147
105
40
C
Adrenal
Cerebellum
CNS E14
CNS E18
Colon
Cortex
Frontal lobe
Heart
Kidney
Large Intestine
Liver Adult
Liver E14
Liver E14.5
Liver E18
Lung
Mammary Gland
Ovary
Spleen
Subcutaneous Fat
Testis
Thymus
Whole Brain E14.5
Gm20939
Gm19705
Xist
Mir17hg
Snhg1
Gas5
Gm20703
Dleu2
6720427I07Rik
Epb41l4aos
Snhg5
Mir142hg
A330094K24Rik
Firre
Dancr
5430416N02Rik
Platr6
D5Ertd605e
B830012L14Rik
Gm5165
Snhg18
Neat1
Bvht
Gm26945
Gm28373
2900041M22Rik
Snhg14
Sptbn5
4930444M15Rik
Mir124a-1hg
Gm12688
Platr15
Platr12
Gm26763
Rbakdn
Gm31693
Gm38509
C330004P14Rik
E130202H07Rik
Gm47599
0
1
D
Relative LINC00862 Levels
**
***
Control
Cirrhotic liver
HCC
E
SNU-182
Huh1
Huh7
JHH2
SNU-387
Hep3B
Alex
HepG2
Li7
THLE-2
Relative LINC00862 Levels

Figure 1 - figure supplement 1







E

GO-Term (biol. process)	P-value (Fisher)
cell cycle	< 1e-30
DNA metabolic process	< 1e-30
cell cycle process	< 1e-30
nucleic acid metabolic process	< 1e-30
cellular macromolecule metabolic proces	< 1e-30
nucleobase-containing compound metab	< 1e-30
heterocycle metabolic process	< 1e-30
mitotic cell cycle	< 1e-30
cellular aromatic compound metabolic pr	< 1e-30
cellular nitrogen compound metabolic pr	< 1e-30
nuclear division	< 1e-30
mitosis	< 1e-30
organic cyclic compound metabolic proce	< 1e-30
organelle fission	< 1e-30
nitrogen compound metabolic process	< 1e-30
cell division	< 1e-30
DNA replication	< 1e-30
macromolecule metabolic process	< 1e-30
cellular metabolic process	< 1e-30
organelle organization	< 1e-30
DNA repair	< 1e-30
primary metabolic process	< 1e-30
chromosome organization	< 1e-30
chromosome segregation	< 1e-30
response to DNA damage stimulus	< 1e-30

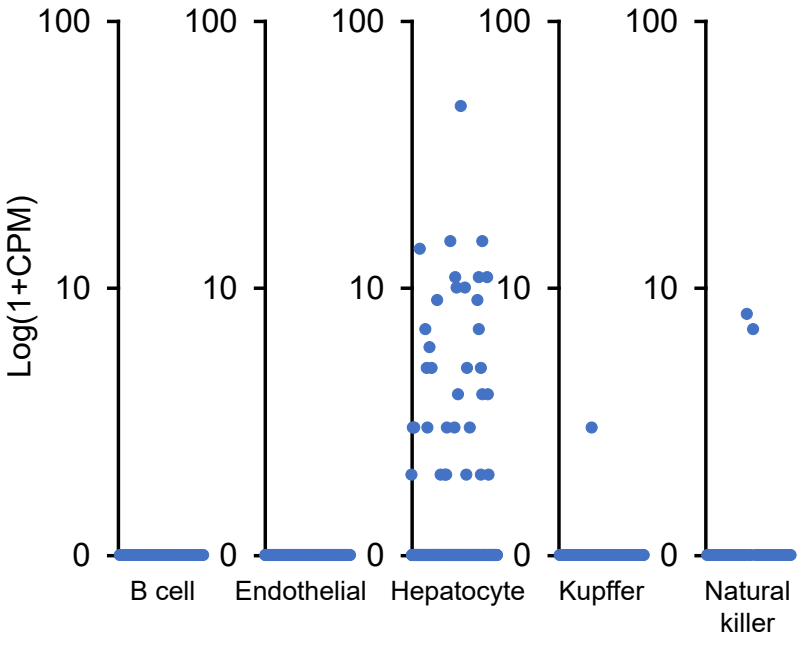


Figure 2

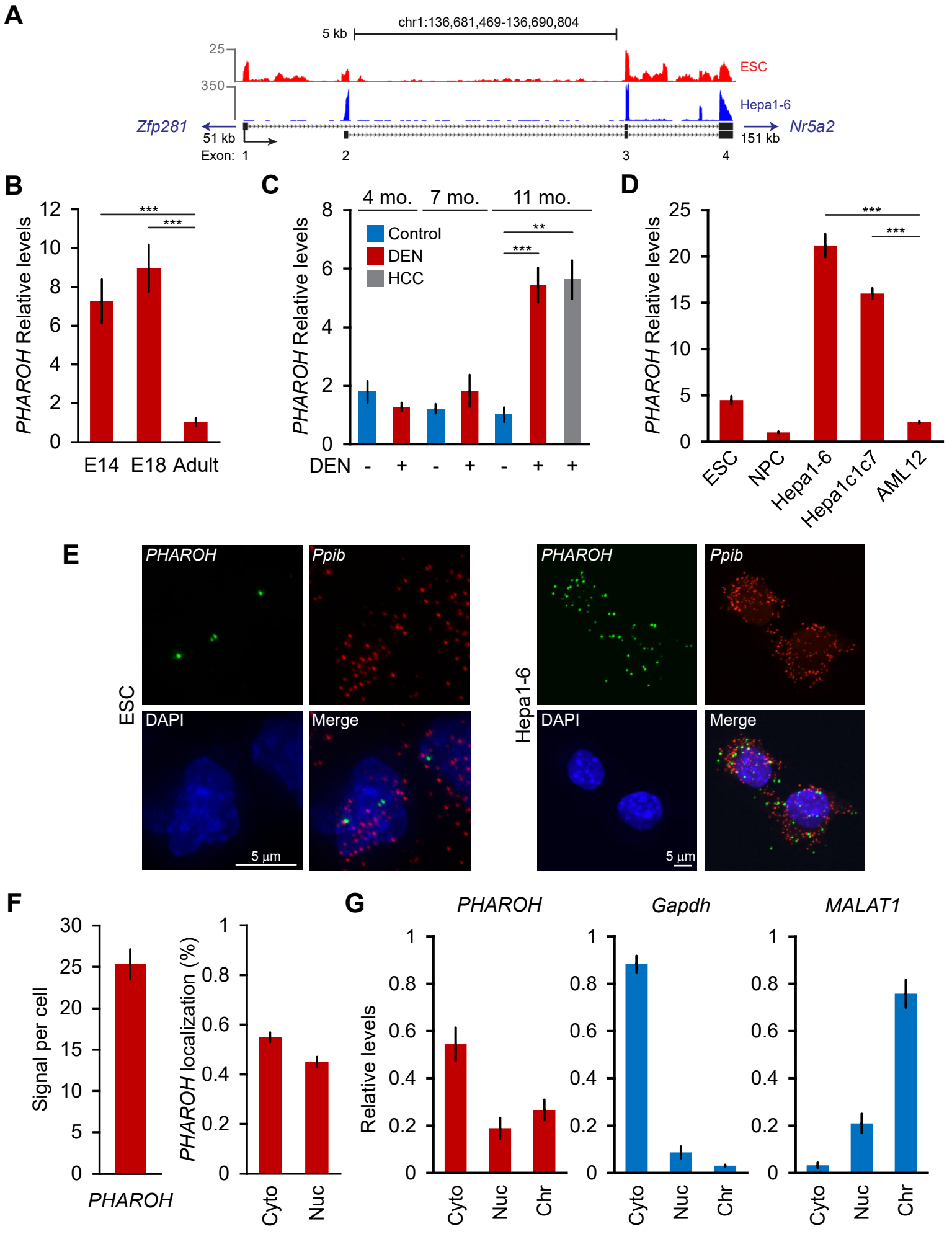

A
chr1:136,681,469-136,690,804
5 kb
25
350
ESC
Hepa1-6
Zfp281
51 kb
Exon: 1
2
3
4
151 kb
Nr5a2
B
PHAROH Relative levels
E14 E18 Adult
C
4 mo. 7 mo. 11 mo.
Control
DEN
HCC
DEN - + - + - + +
D
ESC NPC Hepa1-6 Hepa1c1c7 AML12
E
PHAROH
Ppib
DAPI
Merge
5 μm
ESC
Hepa1-6
F
Signal per cell
PHAROH
PHAROH localization (%)
Cyto
Nuc
G
PHAROH
Gapdh
MALAT1
Relative levels
Cyto
Nuc
Chr

Figure 2 - figure supplement 2

A

	CPC	CPAT
PHAROH	-0.854	0.046
Gapdh	13.614	0.998

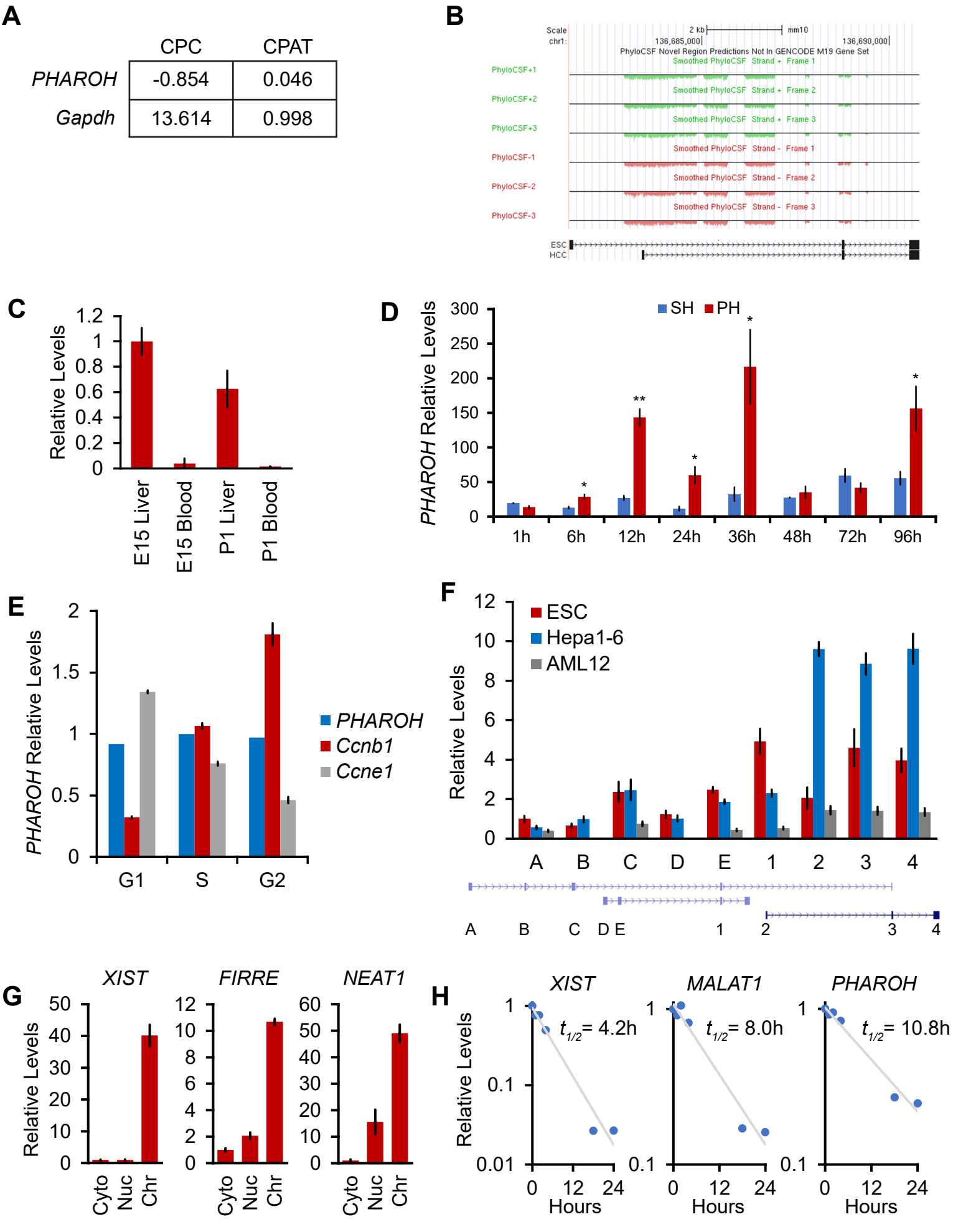


Figure 3

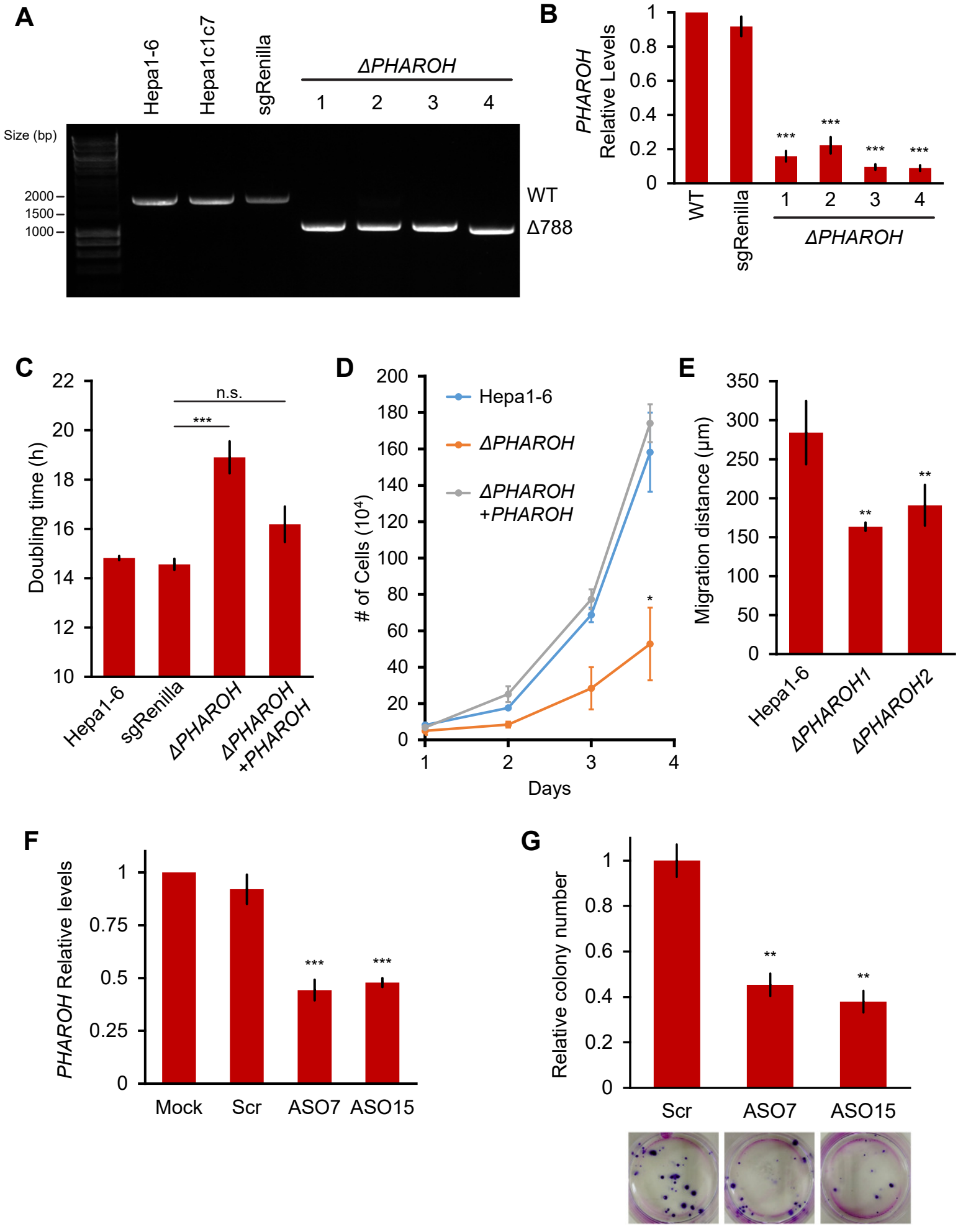

A
Hepa1-6
Hepa1c1c7
sgRenilla
ΔPHAROH
1
2
3
4
Size (bp)
2000
1500
1000
WT
Δ788
B
PHAROH Relative Levels
1
0.8
0.6
0.4
0.2
0
***
WT
sgRenilla
1
2
3
4
ΔPHAROH
C
Doubling time (h)
22
20
18
16
14
12
10
n.s.
***
Hepa1-6
sgRenilla
ΔPHAROH
ΔPHAROH +PHAROH
D
# of Cells (10⁴)
200
180
160
140
120
100
80
60
40
20
0
Hepa1-6
ΔPHAROH
ΔPHAROH +PHAROH
*
1
2
3
4
Days
E
Migration distance (μm)
350
300
250
200
150
100
50
0
**
**
Hepa1-6
ΔPHAROH1
ΔPHAROH2
F
PHAROH Relative levels
1
0.75
0.5
0.25
0
***
***
Mock
Scr
ASO7
ASO15
G
Relative colony number
1
0.8
0.6
0.4
0.2
0
**
**
Scr
ASO7
ASO15

Figure 3 - figure supplement 3

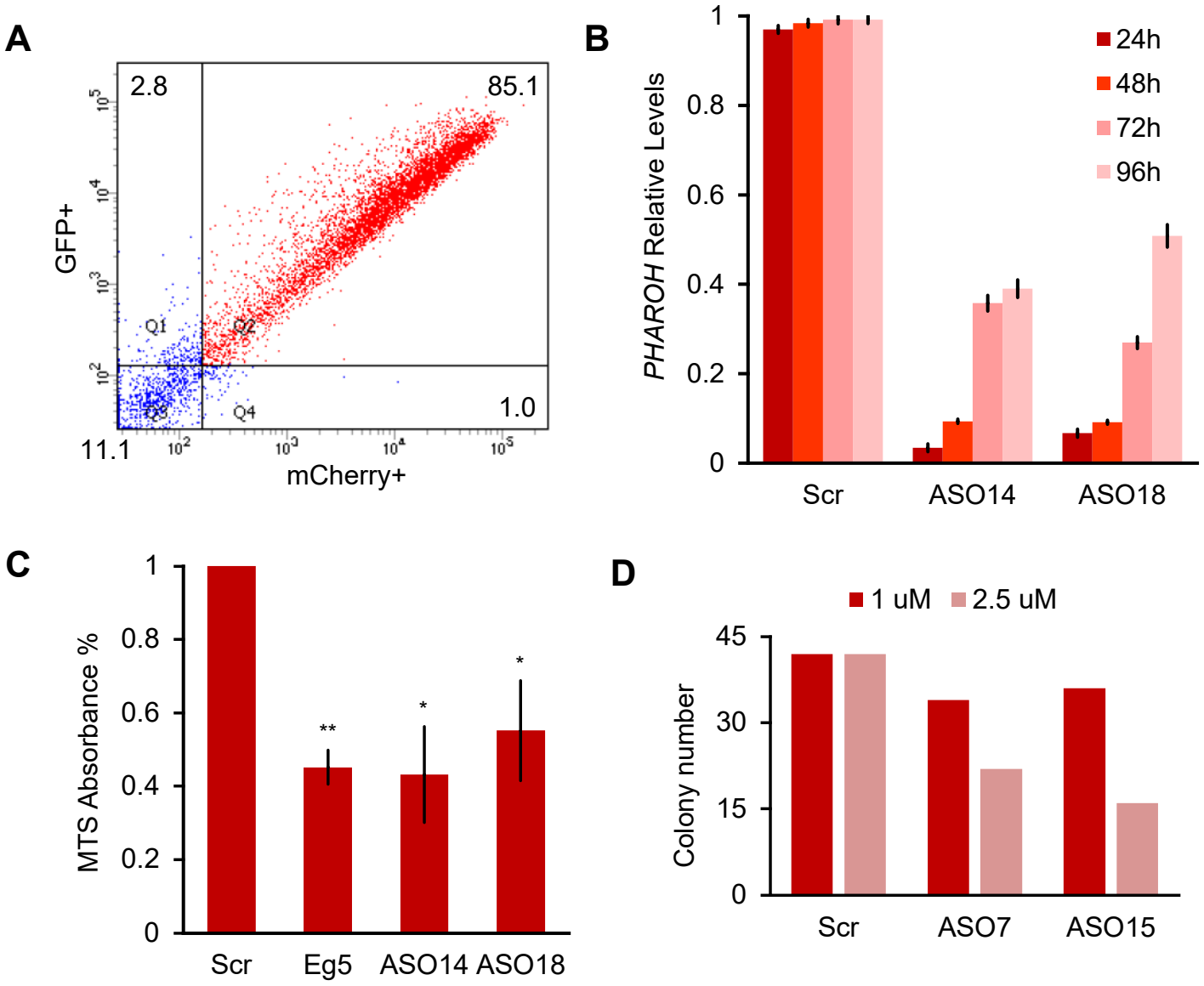
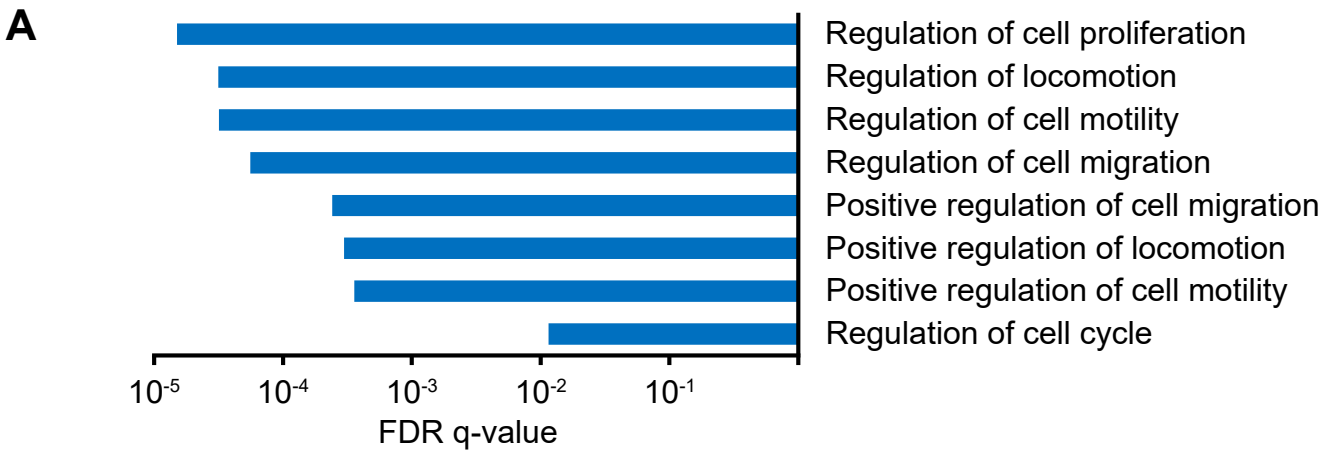
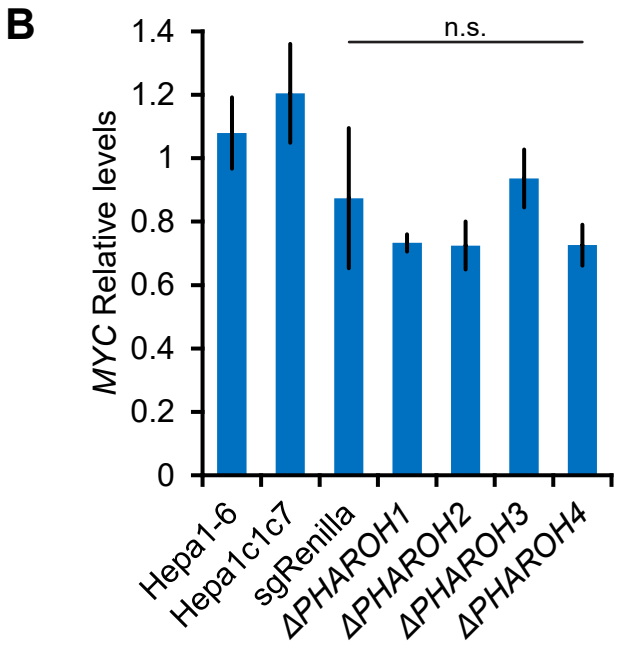
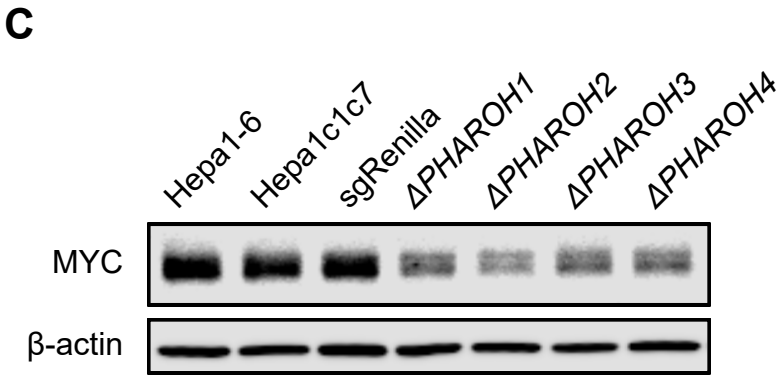


Figure 4


A
Regulation of cell proliferation
Regulation of locomotion
Regulation of cell motility
Regulation of cell migration
Positive regulation of cell migration
Positive regulation of locomotion
Positive regulation of cell motility
Regulation of cell cycle
10-5
10-4
10-3
10-2
10-1
FDR q-value


B
n.s.
MYC Relative levels
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Hepa1-6
Hepa1c1c7
sgRenilla
ΔPHAROH1
ΔPHAROH2
ΔPHAROH3
ΔPHAROH4


C
Hepa1-6
Hepa1c1c7
sgRenilla
ΔPHAROH1
ΔPHAROH2
ΔPHAROH3
ΔPHAROH4
MYC
β-actin

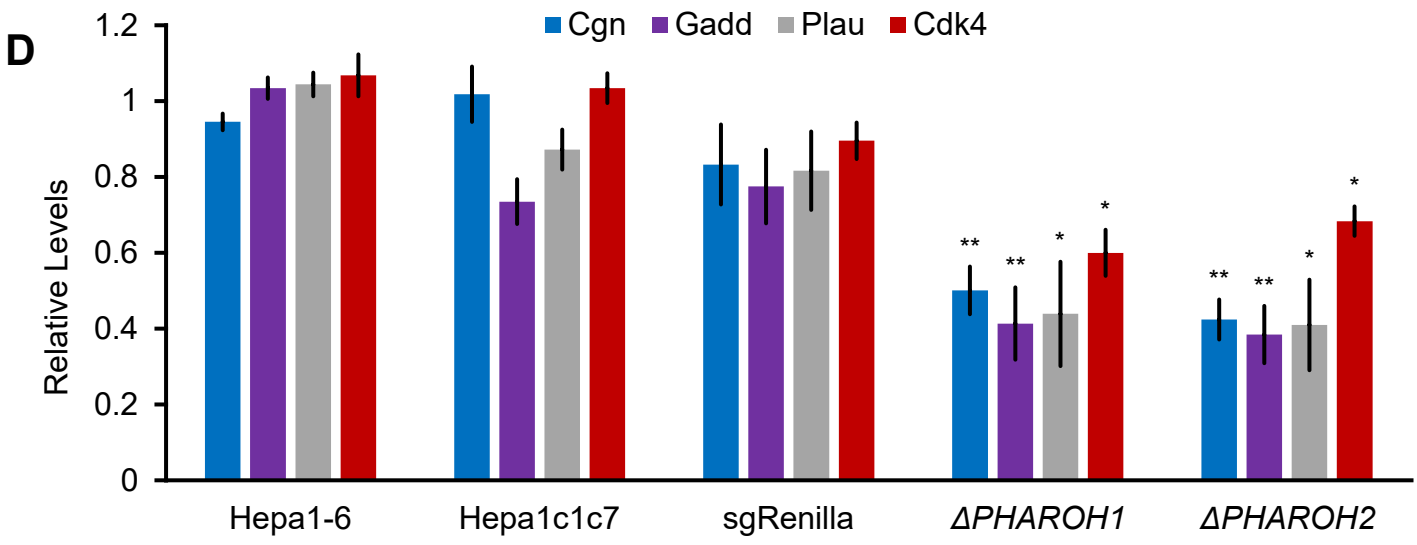
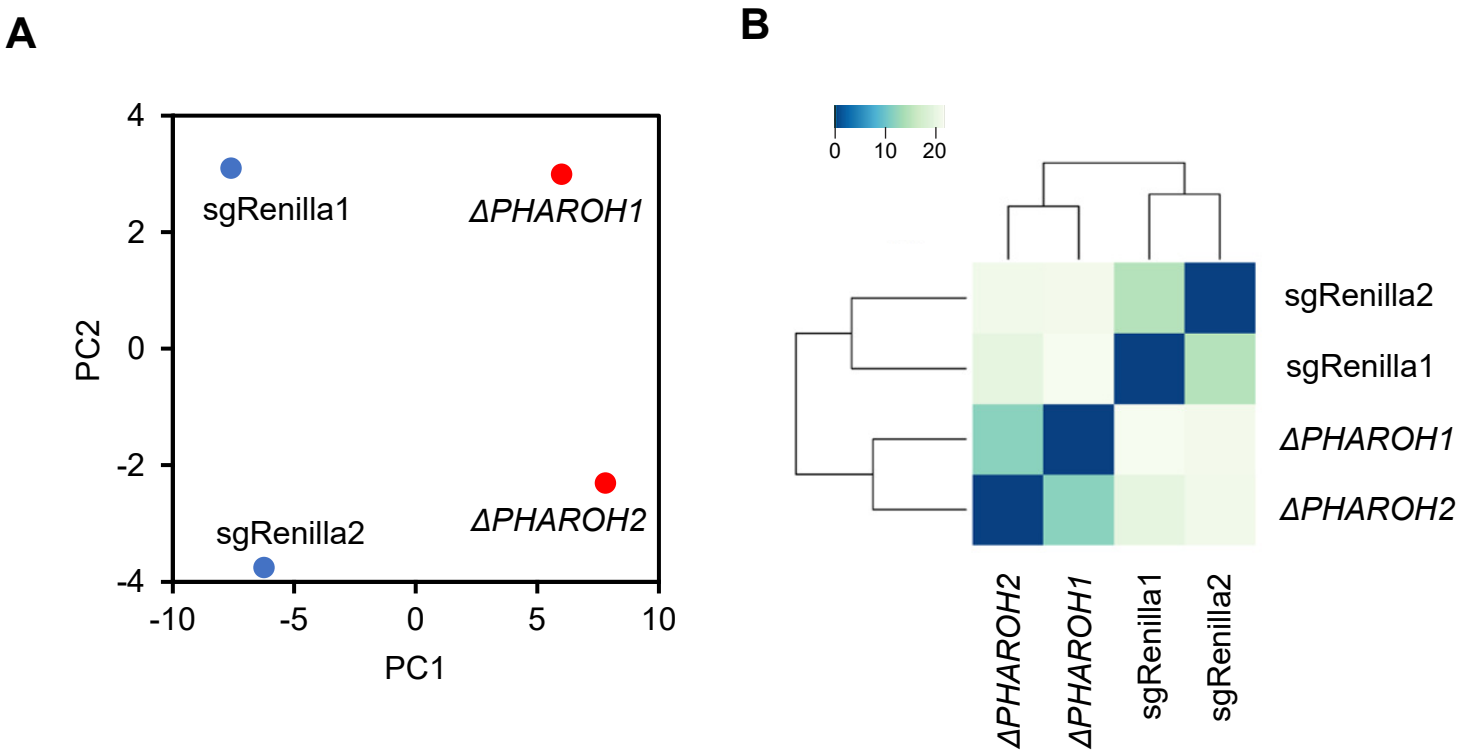

D
Cgn
Gadd
Plau
Cdk4
Relative Levels
1.2
1
0.8
0.6
0.4
0.2
0
Hepa1-6
Hepa1c1c7
sgRenilla
ΔPHAROH1
ΔPHAROH2

Figure 4 - figure supplement 4



C

	log p-value	% Targets	% Background	Enrichment
Myc(bHLH)	-5.47	0.192	0.1551	1.238

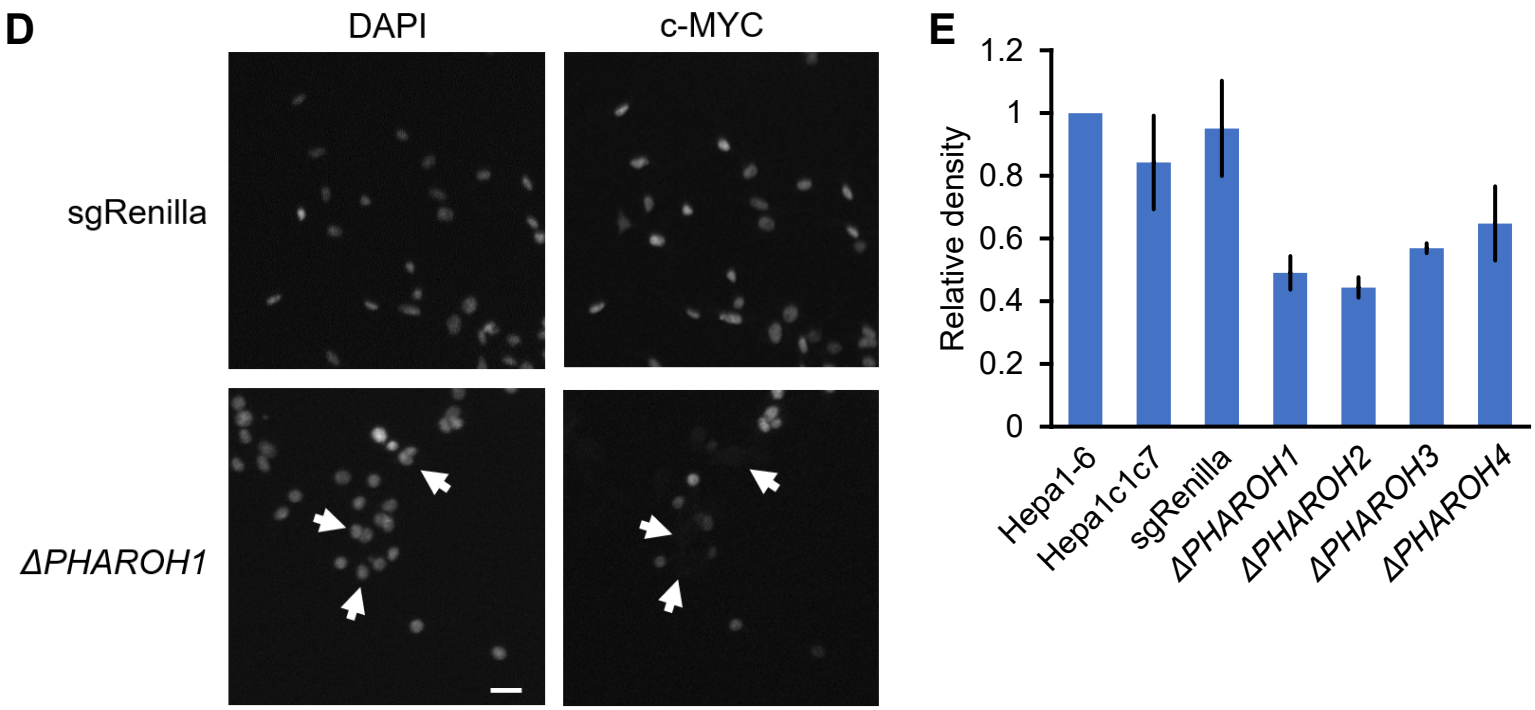


Figure 5.

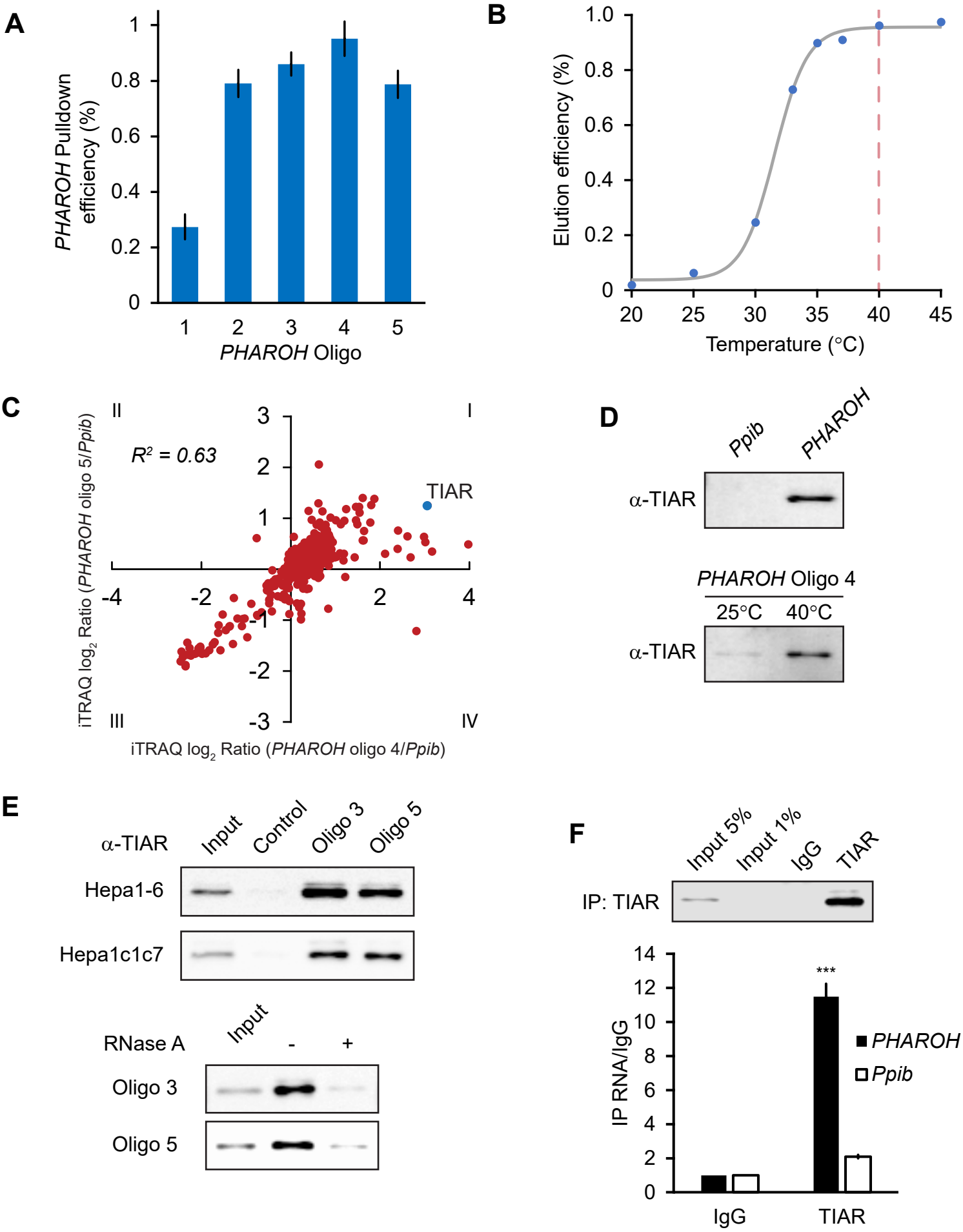

A
PHAROH Pulldown efficiency (%)
PHAROH Oligo
B
Elution efficiency (%)
Temperature (°C)
C
R² = 0.63
TIAR
iTRAQ log2 Ratio (PHAROH oligo 5/Ppib)
iTRAQ log2 Ratio (PHAROH oligo 4/Ppib)
D
Ppib
PHAROH
α-TIAR
PHAROH Oligo 4
25°C
40°C
E
α-TIAR
Input
Control
Oligo 3
Oligo 5
Hepa1-6
Hepa1c1c7
RNase A
F
Input 5%
Input 1%
IgG
TIAR
IP: TIAR
IP RNA/IgG
PHAROH
Ppib

Figure 5 - figure supplement 5

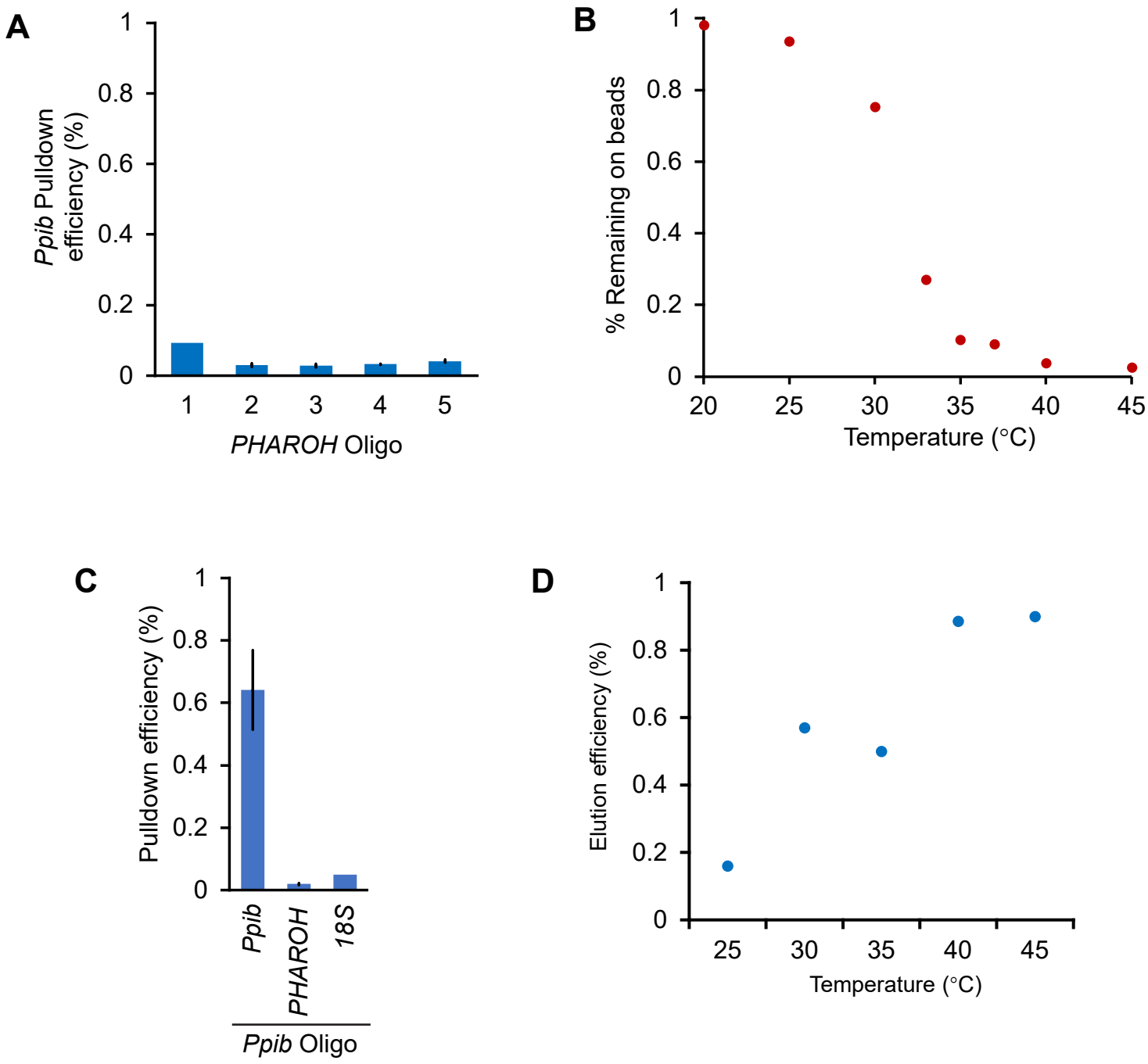


Figure 6.

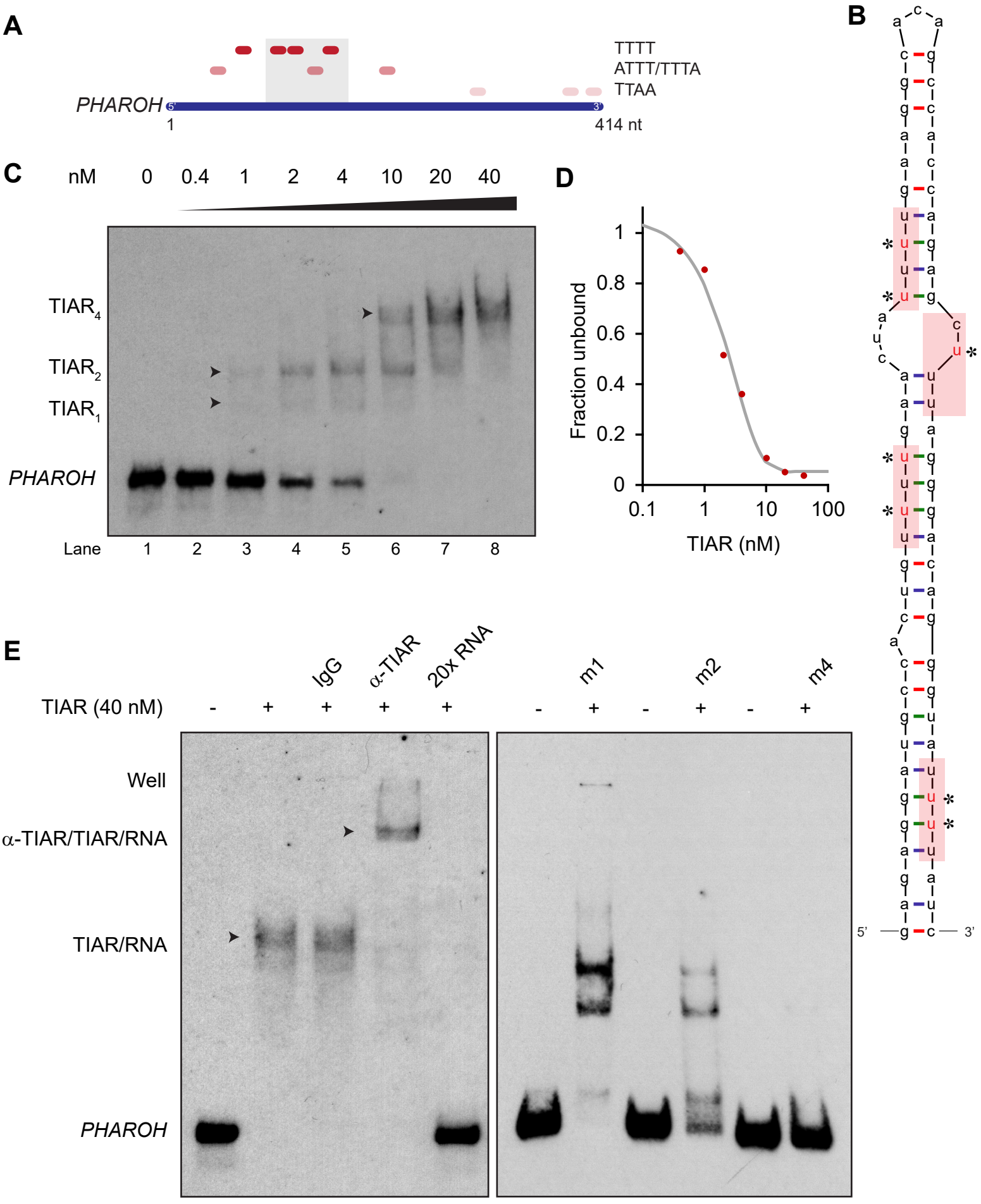
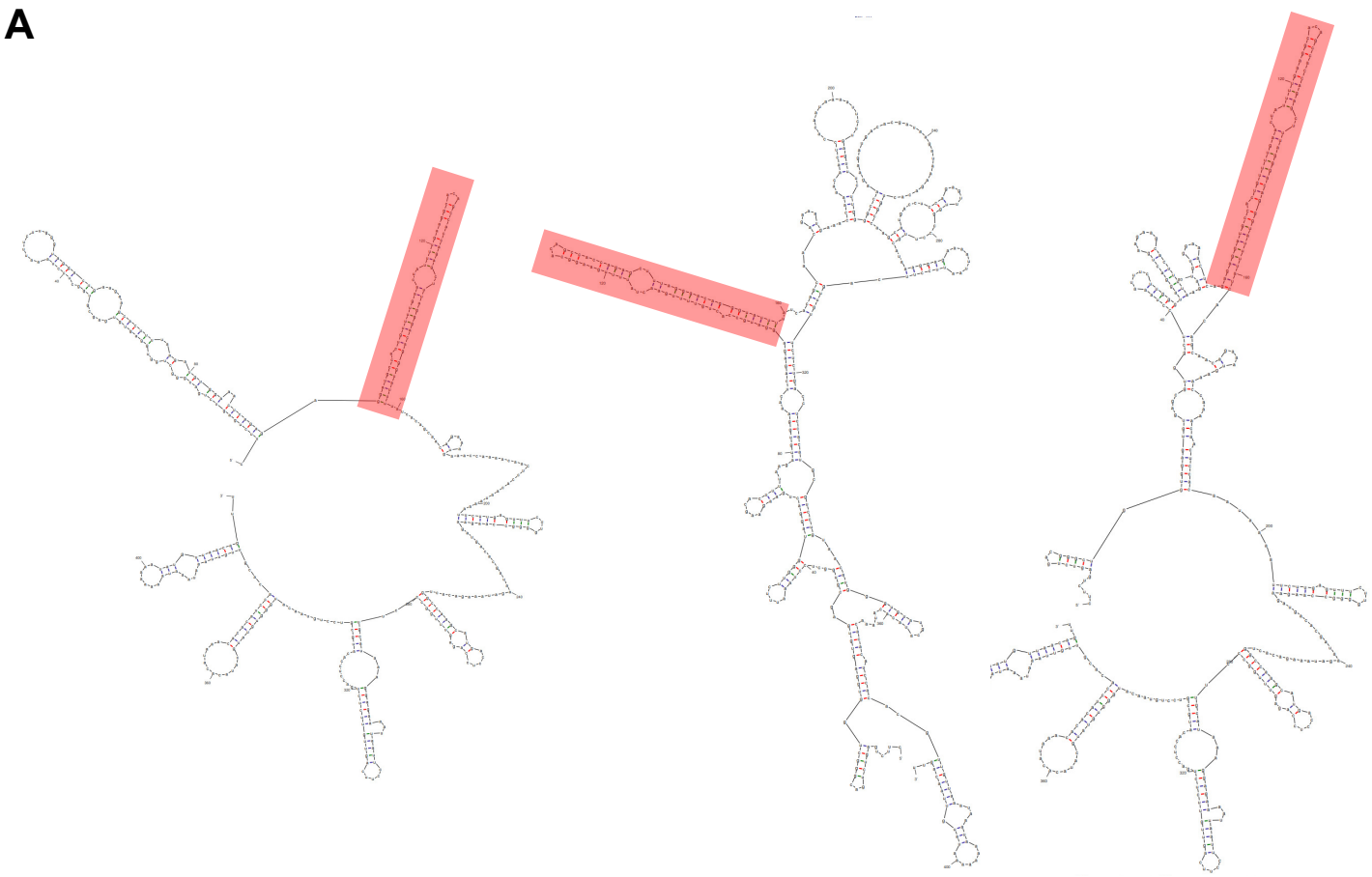
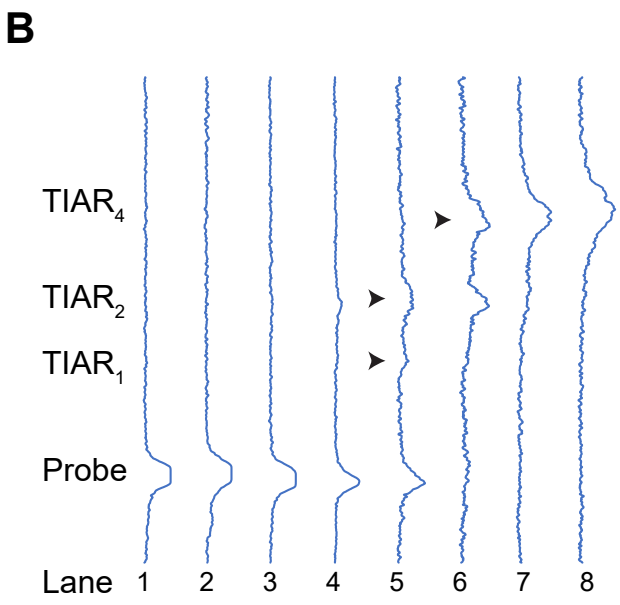

A
PHAROH
5'
3'
1
414 nt
TTTT
ATTT/TTTA
TTAA
B
C
nM 0 0.4 1 2 4 10 20 40
TIAR4
TIAR2
TIAR1
PHAROH
Lane 1 2 3 4 5 6 7 8
D
Fraction unbound
1
0.8
0.6
0.4
0.2
0
0.1 1 10 100
TIAR (nM)
E
IgG
α-TIAR
20x RNA
m1
m2
m4
TIAR (40 nM) - + + + + - + - + - +
Well
α-TIAR/TIAR/RNA
TIAR/RNA
PHAROH

Figure 6 - figure supplement 6





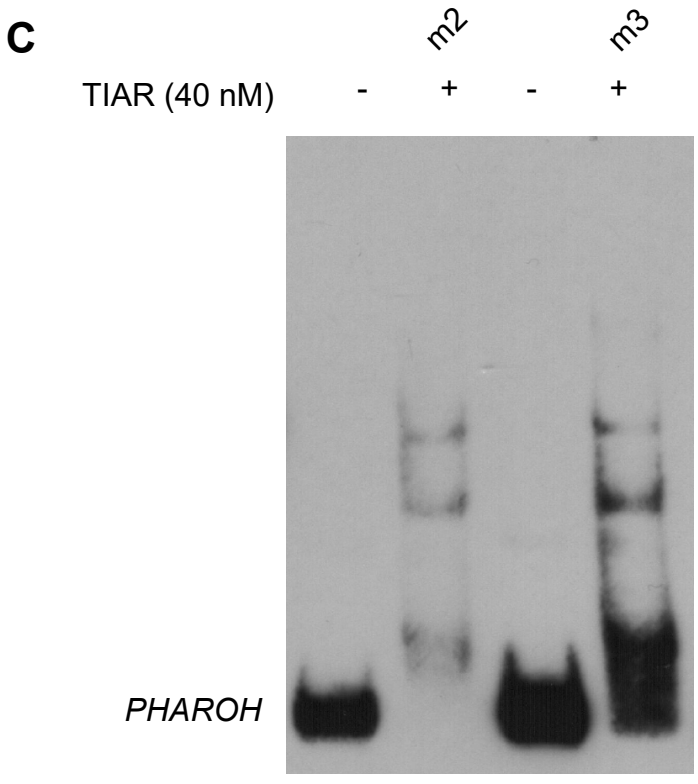


Figure 7.

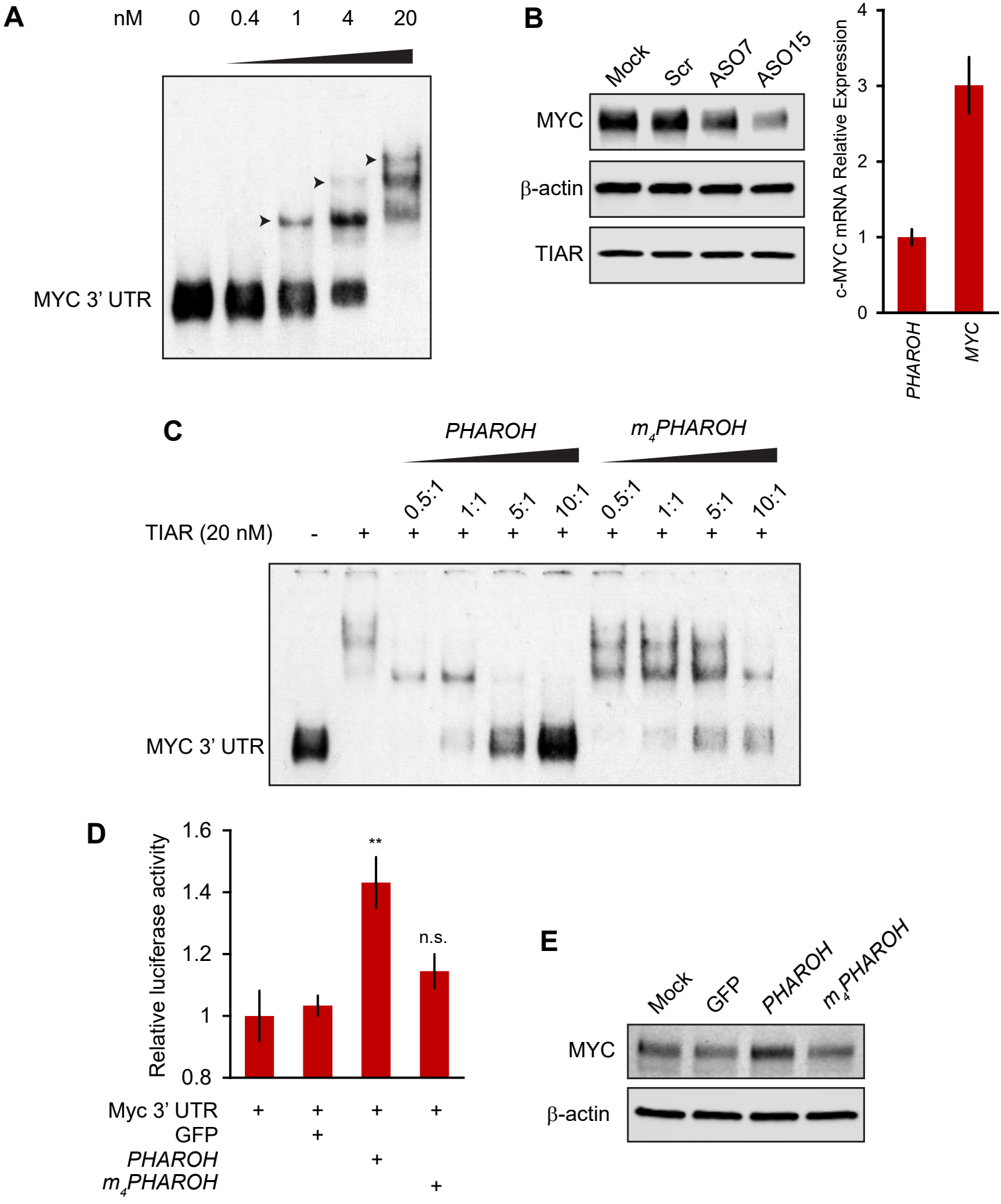

A
nM 0 0.4 1 4 20
MYC 3' UTR
B
Mock Scr ASO7 ASO15
MYC
β-actin
TIAR
c-MYC mRNA Relative Expression
PHAROH
MYC
C
PHAROH
m4PHAROH
0.5:1 1:1 5:1 10:1 0.5:1 1:1 5:1 10:1
TIAR (20 nM) - +
MYC 3' UTR
D
Relative luciferase activity
1.6 1.4 1.2 1 0.8
**
n.s.
Myc 3' UTR + + + +
GFP +
PHAROH +
m4PHAROH +
E
Mock GFP PHAROH m4PHAROH
MYC
β-actin

Figure 7 - figure supplement 7